

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2004, 15:56:05 ; Search time 3413 Seconds
(without alignments)
9921.979 Million cell updates/sec

Title: US-10-091-628-1
Perfect score: 1134
Sequence: 1 atgagagccaattgttccag.....acatcacttcatgtgaatag 1134

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*

28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result			%		Query				ID	Description
			No.	Score	Match	Length	DB			
			1	704.4	62.1	2125	11	AK018423		AK018423 Mus muscu
			2	450.4	39.7	652	10	BB613812		BB613812 BB613812
			3	416.2	36.7	952	13	BY720552		BY720552 BY720552
			4	335.4	29.6	972	12	BG872314		BG872314 602790977
			5	327.4	28.9	356	10	BE181226		BE181226 CM2-HT063
			6	307.8	27.1	1047	29	AY413909		AY413909 Mus muscu
			7	297.8	26.3	1047	29	AY413907		AY413907 Homo sapi
			8	282.8	24.9	773	14	CF998755		CF998755 AGENCOURT
			9	265.4	23.4	364	13	BY779230		BY779230 BY779230
			10	256.4	22.6	834	14	CB320835		CB320835 AGENCOURT
			11	245	21.6	912	29	AY413908		AY413908 Pan trogl
			12	240.6	21.2	372	13	BY134433		BY134433 BY134433
			13	227.6	20.1	620	14	CA353647		CA353647 625196 NC
			14	227.2	20.0	378	13	BY135403		BY135403 BY135403
			15	225.2	19.9	823	14	CA496399		CA496399 AGENCOURT
			16	219.2	19.3	347	13	BY333859		BY333859 BY333859
			17	214	18.9	876	14	CK018712		CK018712 AGENCOURT
			18	208.6	18.4	813	14	CK239397		CK239397 AGENCOURT
			19	206.2	18.2	326	13	BY330845		BY330845 BY330845
			20	206.2	18.2	326	13	BY331654		BY331654 BY331654
			21	203.2	17.9	770	14	CK029723		CK029723 AGENCOURT
			22	202.2	17.8	760	14	CF999869		CF999869 AGENCOURT
			23	201.8	17.8	832	14	CA470489		CA470489 AGENCOURT
			24	200	17.6	658	10	BB625035		BB625035 BB625035
			25	200	17.6	662	10	BB664585		BB664585 BB664585
			26	195.6	17.2	332	13	BY151384		BY151384 BY151384
			27	194.2	17.1	642	9	AU176916		AU176916 AU176916
			28	189	16.7	760	12	BI768670		BI768670 603057222
c			29	184	16.2	274	10	BE151388		BE151388 CM2-HT028
c			30	182.6	16.1	543	28	AQ013853		AQ013853 RPCI11-24
c			31	168.2	14.8	485	29	CE076918		CE076918 tigr-gss-
			32	164.6	14.5	538	10	BE030429		BE030429 128499 MA
			33	162.4	14.3	686	9	AW107022		AW107022 um18f05.y
			34	155.2	13.7	333	13	BY345750		BY345750 BY345750
			35	149.4	13.2	666	10	BB625042		BB625042 BB625042
			36	142.2	12.5	919	12	BI149148		BI149148 602912378
c			37	141.8	12.5	660	14	CB443326		CB443326 694201 MA
			38	141.2	12.5	876	10	BF234728		BF234728 602028555
c			39	140.6	12.4	667	14	CB530259		CB530259 737042 MA
			40	140.4	12.4	307	10	BB605283		BB605283 BB605283
			41	140.4	12.4	419	10	BB846568		BB846568 BB846568
			42	138.8	12.2	315	13	BY347079		BY347079 BY347079
			43	138	12.2	628	10	BE031975		BE031975 130763 MA
			44	137.6	12.1	803	9	AI528386		AI528386 ui96g12.y
			45	137.4	12.1	928	10	BF236073		BF236073 602027173

ALIGNMENTS

RESULT 1

AK018423

LOCUS AK018423 2125 bp mRNA linear HTC 20-SEP-2003

DEFINITION Mus musculus 16 days embryo lung cDNA, RIKEN full-length enriched library, clone:8430417G17 product:hypothetical Sodium bile acid symporter containing protein, full insert sequence.

ACCESSION AK018423

VERSION AK018423.1 GI:12858114

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1

AUTHORS Carninci, P. and Hayashizaki, Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253

PUBMED 10349636

REFERENCE 2

AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE 20499374

PUBMED 11042159

REFERENCE 3

AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

JOURNAL Genome Res. 10 (11), 1757-1771 (2000)

MEDLINE 20530913

PUBMED 11076861

REFERENCE 4

AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL Nature 409, 685-690 (2001)

REFERENCE 5

AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

REFERENCE 6 (bases 1 to 2125)

AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

TITLE Direct Submission

JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cdNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.2. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.

FEATURES

source Location/Qualifiers

1. .2125

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="FANTOM_DB:8430417G17"

/db_xref="MGI:1909149"

/db_xref="taxon:10090"

/clone="8430417G17"

/tissue_type="lung"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="16 days embryo"

CDS

173. .1294

/note="unnamed protein product; hypothetical Sodium bile acid symporter containing protein (InterPro|IPR002657, evidence: InterPro)

putative"

/codon_start=1

/protein_id="BAB31203.1"

/db_xref="GI:12858115"

/translation="MSTDCAGNSTCPVNSTEEDPPVGMEGHANLKLLEFTVLSAVMVGL
 VMFSFGCSVESQKLWLHLRRPWGIAVGLLSQFGLMPLTAYLLAIGFGLKPFQAIIVLM
 MGSCPGGTISNVLTFWVDGMDLSISMTTCSTVAALGMMPLCLYIYTRSWTLTQNLVI
 PYQSIGITLVSLVVPVASGVYVNYRWPQATVILKVGAILGGMLLLVAVTGMVLAKG
 WNTDVTLLVISCFPLVGHVTGFLLAFLTHQSWQRCRTISIIETGAQNIQLCIAMLQLS
 FSAEYLVQLNLFALAYGLFQVLHGLLIVAAYQAYKRRQKSKCRRQHPDCPDVCYEKQP
 RETSAFLDKGDEAAVTLGPVQPEQHHRAAELTSHIPSCE"

ORIGIN

Query Match 62.1%; Score 704.4; DB 11; Length 2125;
 Best Local Similarity 77.7%; Pred. No. 6.1e-183;
 Matches 881; Conservative 0; Mismatches 241; Indels 12; Gaps 2;

Qy	1	ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG	60
Db	173	ATGAGCACAGACTGTGCGGGCAACTCCACCTGCCCTGTCAACAGTACGGAGGAAGACCCG	232
Qy	61	CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTCACAGTGGTGTCCACTGTG	120
Db	233	CCCGTGGGAATGGAGGGCCATGCGAATCTAAAGCTGCTTTTACAGTGCTCTCGGCTGTG	292
Qy	121	ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCTG	180
Db	293	ATGGTGGGTTTGGTCATGTTCTCTTTGGATGTTCTGTGGAGAGTCAGAAGCTCTGGTTG	352
Qy	181	CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT	240
Db	353	CACCTCAGAAGACCCTGGGGCATCGCAGTGGGCCTGCTTCCCAGTTTGGACTTATGCCT	412
Qy	241	TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT	300
Db	413	CTGACAGCTTATCTGTTAGCCATTGGCTTCGGTCTGAAACCATTCCAAGCTATTGCTGTC	472
Qy	301	CTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGAT	360
Db	473	CTCATGATGGGGAGCTGCCCTGGGGGGCACCATCTCTAATGTTCTCACCTTCTGGGTTGAT	532
Qy	361	GGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATG	420
Db	533	GGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACAGTGGCCGCCCTGGGAATG	592
Qy	421	ATGCCACTCTGCATTTATCTCTACACCTGGTCCCTGGAGTCTTCAGCAGAATCTCACCATT	480
Db	593	ATGCCTCTCTGCCTCTACATCTACACCGGTCCTGGACTCTGACACAGAACCTCGTCATT	652
Qy	481	CCTTATCAGAACATAGGAATTACCCTTGTGTGCCCTGACCATTCCTGTGGCCTTTGGTGTC	540
Db	653	CCGTATCAGAGCATAGGAATTACCCTTGTGTCCCTGGTGGTTCCTGTGGCTTCTGGCGTC	712
Qy	541	TATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTT	600
Db	713	TATGTGAATTATAGGTGGCCAAAGCAAGCAACGGTCATTCTCAAGGTCGGAGCCATTCTG	772
Qy	601	GGTGGGGTCCCTCCTTCTGGTGGTGCAGTTGCTGGTGTGGTCCCTGGCGAAAGGATCTTGG	660
Db	773	GGTGGCATGCTCCTCCTGGTGGTGGCAGTTACTGGCATGGTCCCTGGCAAAGG---CTGG	829

Qy	661	AATTCAGACATCACCCCTTCTGACCATCAGTTTCATCTTTCCCTTTGATTGGCCATGTCACG	720
Db	830	AACACAGACGTCACCTCTTCTGGTCATCAGCTGCATTTTCCCTTGGTCGGCCATGTCACA	889
Qy	721	GGTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCCCTTA	780
Db	890	GGCTTCCTGCTGGCATTCTCACCACCAATCTTGGCAAAGGTGCAGGACCATTTCCTATA	949
Qy	781	GAAACTGGAGCTCAGAATATTCAGATGTGCATCACCATGCTCCAGTTATCTTTCAGTGT	840
Db	950	GAGACTGGCGCTCAGAACATCCAGCTGTGCATCGCCATGCTGCAGCTGTCTTCTCTGCT	1009
Qy	841	GAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGAT	900
Db	1010	GAGTACCTGGTCCAGCTGCTAAACTTTGCATTGGCCTATGGACTCTTCCAAGTGTCTGCAC	1069
Qy	901	GGATTTCTTATTGTTGCAGCATATCAGACGTACAAGAGGAGATTGAAGAACAAACATGGA	960
Db	1070	GGGCTGCTCATTTGTCGCAGCATATCAGGCATACAAGAGGAGGCAGAAGAGTAAATGCAGG	1129
Qy	961	AAAAAGAACTCAGGTTGCACAGAAGTCTGCCATACGAGGAAATCGACTTCTTCCAGAGAG	1020
Db	1130	AGACAGCACCCGGATTGCCAGACGTCTGCTACGAGAAGCA-----GCCAGAGAG	1180
Qy	1021	ACCAATGCCTTCTTGGAGGTGAATGAAGAAGGTGCCATCACTCCTGGGCCACCAGGGCCA	1080
Db	1181	ACCAGTGCTTTCTTGGATAAAGGGGATGAGGCTGCCGTAACCTCTGGGGCCAGTGCAGCCA	1240
Qy	1081	ATGGATTGCCACAGGGCTCTCGAGCCAGTTGGCCACATCACTTCATGTGAATAG	1134
Db	1241	GAGCAGCACCCAGGGCTGCTGAGCTGACTAGCCACATTCTTCATGTGAATAG	1294

RESULT 2

BB613812

ACCESSION BB613812

KEYWORDS EST.

ORGANISM *Mus musculus*

REFERENCE 1 (bases 1 to 652)

TITLE BIKEN Mouse ESTs (Arakawa.T., et al., 2001)

COMMENT Contact: Yoshihide Hayashizaki

Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. *Genome Res.* . 10 (10), 1617-1630 (2000)

wagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,
Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.
and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* .
10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,
Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. *Genome Res.* . 11 (2), 281-289 (2001)

Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I.,
Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and
Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences. *Mamm. Genome.* 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

e mouse tissues.

FEATURES

source

Location/Qualifiers

1. .652

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="4831431E11"

/sex="mixed"

/tissue_type="head"

/dev_stage="0 day neonate"

/lab_host="DH10B"

/clone_lib="RIKEN full-length enriched, 0 day neonate
head"

/note="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 100.0. Second

strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN

```

Query Match          39.7%;  Score 450.4;  DB 10;  Length 652;
Best Local Similarity 81.7%;  Pred. No. 5.7e-113;
Matches 533;  Conservative 0;  Mismatches 116;  Indels 3;  Gaps 1;

Qy      92 AGCTCGTTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGTTCTCTTTGGGAT 151
      ||||  ||||  ||||  ||  ||  ||||  ||||  ||  ||  ||||  ||||  ||||
Db      4  AGCTGCTTTTTACAGTGCTCTCGGCTGTGATGGTGGGTTTGGTCATGTTCTCTTTGGAT 63

Qy     152 GTTCCGTGGAGATCCGGAAGCTGTGGTCGCACATCAGGAGACCCTGGGGCATTGCTGTGG 211
      ||||  ||||  ||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
Db      64 GTTCTGTGGAGAGTCAGAAGCTCTGGTTGCACCTCAGAAGACCCTGGGGCATCGCAGTGG 123

Qy     212 GACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGGCCATTAGCTTTT 271
      |  ||||  |  ||||  ||||  ||  ||||  |  ||||  ||||  ||||  ||||
Db     124 GCCTGCTTTCCCAGTTTGGACTTATGCCTCTGACAGCTTATCTGTTAGCCATTGGCTTCG 183

Qy     272 CTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCCCCGGGGGGCACCA 331
      ||||  |||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
Db     184 GTCTGAAACCATTCCAAGCTATTGCTGTCCTCATGATGGGGAGCTGCCCTGGGGGCACCA 243

Qy     332 TCTCTAACATTTTCACCTTCTGGGTTGATGGAGATATGGATCTCAGCATCAGTATGACAA 391
      ||||  ||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
Db     244 TCTCTAATGTTCTCACCTTCTGGGTTGATGGAGATATGGATCTCAGCATCAGTATGACAA 303

Qy     392 CCTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATCTCTACACCTGGT 451
      ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
Db     304 CCTGTTCCACAGTGGCCGCCCTGGGAATGATGCCTCTCTGCCTCTACATCTACACCCGGT 363

Qy     452 CCTGGAGTCTTCAGCAGAATCTCACCATTTCCTTATCAGAACATAGGAATTACCCTTGTGT 511
      ||||  |||  ||||  |||  ||||  ||||  ||||  ||||  ||||  ||||
Db     364 CCTGGACTCTGACACAGAACCTCGTCATTCCGTATCAGAGCATAGGAATTACCCTTGTGT 423

Qy     512 GCCTGACCATTCTGTGGCCTTTGGTGTCTATGTGAATTACAGATGGCCAAAACAATCCA 571
      ||||  ||||  ||||  ||  ||||  ||||  ||||  ||||  ||||  ||||
Db     424 CCCTGGTGGTTCTGTGGCTTCTGGCGTCTATGTGAATTATAGGTGGCCAAAGCAAGCAA 483

Qy     572 AAATCATTCTCAAGATTGGGGCCGTTGTTGGTGGGGTCCTCCTTCTGGTGGTCGCAGTTG 631
      ||||  ||||  ||  ||  ||  ||  ||||  |  ||||  ||||  ||||
Db     484 CGGTCATTCTCAAGGTCGGAGCCATTCTGGGTGGCATGCTCCTCCTGGTGGTGGCAGTTA 543

Qy     632 CTGGTGTGGTCCTGGCGAAAGGATCTTGAATTACAGACATCACCTTCTGACCATCAGTT 691
      ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
Db     544 CTGGCATGGTCCTGGCAAAGG---CTGGAACACAGACGTCACCTTCTGGTCATCAGCT 600

Qy     692 TCATCTTTCCTTTGATTGGCCATGTCACGGGTTTTCTGCTGGCACTTTTAC 743
      |||  ||  ||  |||  |  ||||  ||||  ||  ||  ||||  ||  ||
Db     601 GCATTTTCCCCTTGGTCGGCCATGTCACAGGCTTCCTGCTGGGCATTCTCAC 652

```

RESULT 3

BY720552
 LOCUS BY720552 952 bp mRNA linear EST 17-DEC-2002
 DEFINITION BY720552 RIKEN full-length enriched, 16 days embryo lung Mus
 musculus cDNA clone 8430417G17 5', mRNA sequence.
 ACCESSION BY720552
 VERSION BY720552.1 GI:27133669
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 952)
 AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
 Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
 Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
 Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
 Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
 Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
 Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
 Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
 Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
 Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
 Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
 Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
 Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
 Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
 Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
 Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
 Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
 Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
 Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
 Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
 Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
 Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
 Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
 Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
 Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
 Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
 Rogers,J., Birney,E. and Hayashizaki,Y.
 TITLE Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 MEDLINE 22354683
 PUBMED 12466851
 COMMENT Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp,
 URL:http://genome.gsc.riken.go.jp/
 Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P.,
 Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F.,
 Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y.,
 Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M.,

Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y.
Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

source

Location/Qualifiers

1. .952

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="8430417G17"

/sex="mixed"

/tissue_type="lung"

/dev_stage="16 days embryo"

/lab_host="DH10B"

/clone_lib="RIKEN full-length enriched, 16 days embryo lung"

/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

ORIGIN

Query Match 36.7%; Score 416.2; DB 13; Length 952;
Best Local Similarity 80.9%; Pred. No. 1.9e-103;

Matches 509; Conservative 0; Mismatches 118; Indels 2; Gaps 2;

```
Qy      1 ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG 60
      ||||| | | ||| | ||| ||| ||||| ||||| | ||||| || | |
Db     173 ATGAGCACAGACTGTGCGGGCAACTCCACCTGCCCTGTCAACAGTACGGAGGAAGACCCG 232

Qy     61 CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTG 120
      || ||||| ||||| |||| || || |||| |||| ||||| | || |||||
Db    233 CCCGTGGGAATGGAGGGCCATGCGAATCTAAAGCTGCTTTTTACAGTGCTCTCGGCTGTG 292

Qy    121 ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCG 180
      ||| |||| || ||||| ||||| ||||| ||||| | ||||| |||| |
Db    293 ATGGTGGGTTTGGTCATGTTCTCTTTGGATGTTCTGTGGAGAGTCAGAAGCTCTGGTTG 352

Qy    181 CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT 240
      ||| |||| ||||| ||||| || ||||| ||||| | ||||| || |||||
Db    353 CACCTCAGAAGACCCTGGGGCATCGCAGTGGGCCTGCTTTCCAGTTTGGACTTATGCCT 412

Qy    241 TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT 300
      | ||||| ||||| | ||||| |||| ||||| ||| ||||| ||||| |||||
Db    413 CTGACAGCTTATCTGTTAGCCATTGGCTTCGGTCTGAAACCATTCCAAGCTATTGCTGTC 472

Qy    301 CTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTACCTTCTGGGTTGAT 360
      ||||| ||||| ||||| ||||| ||||| ||||| || ||||| ||||| |||||
Db    473 CTCATGATGGGGAGCTGCCCTGGGGGGCACCATCTCTAATGTTCTCACCTTCTGGGTTGAT 532

Qy    361 GGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATG 420
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    533 GGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACAGTGGCCGCCCTGGGAATG 592

Qy    421 ATGCCACTCTGCATTTATCTCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATT 480
      ||||| ||||| | || ||||| ||||| |||| ||||| |||| |||| ||||
Db    593 ATGCCTCTCTGCCTCTACATCTACACCCGGTCTGGACTCTGACACAGAAACTCGTCATT 652

Qy    481 CCTTATCAGAACATAGGAATTACCCTTGTGTGCCTGACCATTCTGTGGCCTTTGGTGTC 540
      || ||||| ||||| ||||| ||||| |||| ||||| ||||| |||| ||||
Db    653 CCGTATCAGAGCATAGGAATTACCCTTGTGTCCCTGGTGGTTCCTGTGGCTTCTGGCGTC 712

Qy    541 TATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTT 600
      ||||| ||||| | |||| ||||| ||| | ||||| | ||| | ||| |
Db    713 TATGTGAATTATAAGTGGCC-AAACAAGCCACAGTCATTCTCTAAGTCGGAGACATTCTG 771

Qy    601 GGTGGGGTCCTCCTTCTGGTGGTCGCAGT 629
      ||||| | ||| ||||| || ||
Db    772 GGTGGCAT-TTCCTGCTGGTGGTGGCGGT 799
```

RESULT 4

BG872314

LOCUS BG872314 972 bp mRNA linear EST 29-MAY-2001

DEFINITION 602790977F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4922227 5', mRNA sequence.

ACCESSION BG872314

VERSION BG872314.1 GI:14222854

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 972)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM10841 row: j column: 20
 High quality sequence stop: 786.

FEATURES Location/Qualifiers
 source 1. .972
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4922227"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NCI_CGAP_SG2"
 /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:
 NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo
 dT. Average insert size 1.3 kb. Constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Query Match 29.6%; Score 335.4; DB 12; Length 972;
 Best Local Similarity 73.8%; Pred. No. 4e-81;
 Matches 458; Conservative 0; Mismatches 151; Indels 12; Gaps 2;

```

Qy      514 CTGACCATTCTGTGGCCTTTGGTGTCTATGTGAATTACAGATGGCCAAAACAATCCAAA 573
      ||| ||||| ||||| ||| ||||| ||||| || ||||| ||| | |
Db      1 CTGGTGGTTCCTGTGGCTTCTGGCGTCTATGTGAATTATAGGTGGCCAAAGCAAGCAACG 60

Qy      574 ATCATTTCTCAAGATTGGGGCCGTTGTTGGTGGGGTCCTCCTTCTGGTGGTCGCAGTTGCT 633
      ||||| ||||| | || ||| || | |||| | |||| | ||||| ||||| ||
Db      61 GTCATTCTCAAGGTCGGAGCCATTCTGGGTGGCATGCTCCTCCTGGTGGTGGCAGTTACT 120

Qy      634 GGTGTGGTCCTGGCGAAAGGATCTTGAATTTCAGACATCACCTTCTGACCATCAGTTTC 693
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||| |
Db      121 GGCATGGTCCTGGCAAAAGG---CTGGAACACAGACGTCCTCTTCTGGTCATCAGCTGC 177

Qy      694 ATCTTTCCCTTTGATTGGCCATGTCACGGGTTTTCTGCTGGCACTTTTTACCCACCAGTCT 753
      || || || ||| | ||||| ||||| || || ||||| || | ||||| |||
Db      178 ATTTTCCCCTTGGTCGGCCATGTCACAGGCTTCCTGCTGGCATTCTCACCCACCAATCT 237

Qy      754 TGGCAAAGGTGCAGGACAATTTCCCTTAGAAACTGGAGCTCAGAATATTAGATGTGCATC 813
      ||||| ||||| ||||| ||||| ||||| ||||| || || ||||| |||
Db      238 TGGCAAAGGTGCAGGACCATTTCCATAGAGACTGGCGCTCAGAACATCCAGCTGTGCATC 297

Qy      814 ACCATGCTCCAGTTATCTTTCAGTCTGCTGAGCACTTGGTCCAGATGTTGAGTTTCCCACTG 873
  
```


Db 298 GCCATGCTGCAGCTGTCCTTCTCTGCTGAGTACCTGGTCCAGCTGCTAAACTTTGCATTG 357
 Qy 874 GCCTATGGACTCTTCCAGCTGATAGATGGATTTCTTATTGTTGCAGCATATCAGACGTAC 933
 Db 358 GCCTATGGACTCTTCCAAGTGCTGCACGGGCTGCTCATTGTCGCAGCATATCAGGCATAC 417
 Qy 934 AAGAGGAGATTGAAGAACAACATGGAAAAAGAACTCAGGTTGCACAGAAGTCTGCCAT 993
 Db 418 AAGAGGAGGCAGAAGAGTAAATGCAGGAGACAGCACCCGGATTGCCAGACGTCTGCTAC 477
 Qy 994 ACGAGGAAATCGACTTCTTCCAGAGAGACCAATGCCTTCTTGAGGTTGAATGAAGAAGGT 1053
 Db 478 GAGAAGCA-----GCCCAGAGAGACCAGTGCTTTCTTGATAAAGGGGATGAGGCT 528
 Qy 1054 GCCATCACTCCTGGGCCACCAGGGCCAATGGATTGCCACAGGGCTCTCGAGCCAGTTGGC 1113
 Db 529 GCCGTAACCTCTGGGGCCAGTGCAGCCAGAGCAGCACCCAGGGCTGCTGAGCTGACTAGC 588
 Qy 1114 CACATCACTTCATGTGAATAG 1134
 Db 589 CACATTCCTTCATGTGAATAG 609

RESULT 5

BE181226

LOCUS BE181226 356 bp mRNA linear EST 22-JUN-2000

DEFINITION CM2-HT0630-220300-125-f05 HT0630 Homo sapiens cDNA, mRNA sequence.

ACCESSION BE181226

VERSION BE181226.1 GI:8660402

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 356)

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

PUBMED 10737800

COMMENT Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=CM2-HT0630-220
300-125-f05&t3=2000-03-22&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 12

High quality sequence stop: 354.

FEATURES
Location/Qualifiers
source 1. .356
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="HT0630"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN

Query Match 28.9%; Score 327.4; DB 10; Length 356;
Best Local Similarity 97.7%; Pred. No. 4.3e-79;
Matches 343; Conservative 0; Mismatches 6; Indels 2; Gaps 1;

```
Qy      411 CCTGGGAATGATGCCACTCTGCATTTATCTCTACACCTGGTCCTGGAGTCTTCAGCAGAA 470
          |||
Db        6 CCTGGGAATGATGCCACTCTGCATTTATCTCTACACCTGGTCCTGGAGTCTTCAGCAGAA 65

Qy      471 TCTCACCATTCCCTTATCAGAACATAGGAA--TTACCCTTGTGTGCCTGACCATTCCCTGTG 528
          |||
Db        66 TCTCACCATTCCCTTATCAGAACATAGGAAGTTACCCTATGTGTGCCTGACCATTCCCTGTG 125

Qy      529 GCCTTTGGTGTCTATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATT 588
          |||
Db       126 GCCTTAGGTGTCTATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATT 185

Qy      589 GGGGCCGTTGTTGGTGGGGTCCTCCTTCTGGTGGTCGCAGTTGCTGGTGTGGTCCTGGCG 648
          |||
Db       186 GGGGCCGTTGTTGGTGGGGTCCTCCTTCTGGTGGTCGCAGTTGCTGGTGTGGTCCTGGCG 245

Qy      649 AAAGGATCTTGAATTCAGACATCACCCCTTCTGACCATCAGTTTCATCTTTCCTTTGATT 708
          |||
Db       246 AAAGGATCTTGAATTCAGACATCACCCCTTCTGACCATCAGTTTCATCTTTCCTTTGATT 305

Qy      709 GGCCATGTCACGGGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGCCAA 759
          |||
Db       306 GGCCATGTCACGGGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGCCA 356
```

RESULT 6

AY413909

LOCUS AY413909 1047 bp DNA linear GSS 17-DEC-2003

DEFINITION Mus musculus SLC10A2 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.

ACCESSION AY413909

VERSION AY413909.1 GI:39769871

Qy 380 TCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATC 439
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 380 TTAGCATGACCACTTGCTCCACACTGCTTGCCCTTGGGAATGATGCCTCTTTGCCTCTTCG 439

Qy 440 TCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATTTCCTTATCAGAACATAGGAA 499
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 440 TCTACACCAAGATGTGGGTTGACTCGGGAACGATTGTGATTCCCTATGATAGCATTGGTA 499

Qy 500 TTACCCTTGTGTGCCTGACCATTCCTGTGGCCTTTGGTGTCTATGTGAATTACAGATGGC 559
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 500 TTTCTCTGGTTGCTCTTGTATTCCTGTTTCCTTTGGAATGTTTGTAATCACAAATGGC 559

Qy 560 CAAAACAATCCAAAATCATTTCTCAAGATTGGGGCCGTTGTTGGTGGGGTCCTCCTTCTGG 619
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 560 CACAAAAAGCGAAGATTATACTTAAAATTGGATCCATCACAGGTGTAATTCTCATTGTGC 619

Qy 620 TGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGATCTTGAATTCAGACATCACCTTC 679
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 620 TCATAGCTGTGATTGGAGGAATACTGTACCAAAGTGCTGGATCATTGAACCCAACTGT 679

Qy 680 TGACCATCAGTTTCATCTTTCCTTTGATTGGCCATGTCACGGGTTTTCTGCTGGCACTTT 739
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 680 GGATTATAGGAACAATATTCCTTATAGCTGGCTACAGCCTGGGTTTTCTTCCTGGCTAGAC 739

Qy 740 TTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCTTAGAAACTGGAGCTCAGAATA 799
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 740 TAGCTGGTCAACCCTGGTACAGGTGCCGAACAGTAGCCTTGGAAACTGGAATGCAGAACA 799

Qy 800 TTCAGATGTGCATCACCATGCTCCAGTTATCTTTCCTGCTGAGCACTTGGTCCAGATGT 859
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 800 CTCAGCTGTGCTCCACCATTTGTACAGCTCTCCTTCTCCCCCGAGGATCTCAACCTGGTGT 859

Qy 860 TGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGATGGATTTCTTATTGTTGCAG 919
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 860 TCACCTTCCCACTCATCTATACTGTTTTCCAGCTCGTCTTTGCAGCAGTAATATTAGGAA 919

Qy 920 CATATCAGACGTACAAGAGGAGATTGAAGAACAAACATG 958
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 920 TTTATGTCACATACAGGAAATGTTATGGAAAAATGATG 958

RESULT 7

AY413907

LOCUS AY413907 1047 bp DNA linear GSS 17-DEC-2003

DEFINITION Homo sapiens SLC10A2 gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.

ACCESSION AY413907

VERSION AY413907.1 GI:39769869

KEYWORDS GSS.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1047)

AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,

Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
 Adams,M.D. and Cargill,M.
 TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios
 JOURNAL Science 302 (5652), 1960-1963 (2003)
 PUBMED 14671302
 REFERENCE 2 (bases 1 to 1047)
 AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
 Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
 Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
 Adams,M.D. and Cargill,M.
 TITLE Direct Submission
 JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 COMMENT This sequence was made by sequencing genomic exons and ordering
 them based on alignment.
 FEATURES Location/Qualifiers
 source 1. .1047
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 gene <1..>1047
 /gene="SLC10A2"
 /locus_tag="HCM5047"
 ORIGIN

Query Match 26.3%; Score 297.8; DB 29; Length 1047;
 Best Local Similarity 58.5%; Pred. No. 9.9e-71;
 Matches 518; Conservative 0; Mismatches 367; Indels 0; Gaps 0;

```

Qy      80 ATGGAAACCTGGAGCTCGTTTTACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGT 139
      || | ||| | || | | ||| || | || ||| || | ||||
Db      80 ATAACATCCTAAGTGTGGTCCTAAGTACGGTGCTGACCATCCTGTTGGCCTTGGTGATGT 139

Qy     140 TCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTTCGCACATCAGGAGACCCTGGG 199
      |||| ||||| ||||| ||| ||| | ||||| || | || ||||
Db     140 TCTCCATGGGATGCAACGTGGAATCAAGAAATTTCTAGGGCACATAAAGCGGCCGTGGG 199

Qy     200 GCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGG 259
      ||||| ||| || | |||| ||||| ||||| | |||| | ||||
Db     200 GCATTTGTGTTGGCTTCCTCTGTTCAGTTTGAATCATGCCCTCACAGGATTCATCCTGT 259

Qy     260 CCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCC 319
      | | |||| | || |||| || | | || |||| || || |||||
Db     260 CGGTGGCCTTTGACATCCTCCCGCTCCAGGCCGTAGTGGTGCTCATTATAGGATGCTGCC 319

Qy     320 CGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGATGGAGATATGGATCTCAGCA 379
      | || || || || || || || || || || || || || || || ||
Db     320 CTGGAGGAAGTGCCTCCAATATCTTGGCCTATTGGGTCGATGGCGACATGGACCTGAGCG 379

Qy     380 TCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATC 439
      |||| |||| || || |||| || || |||| ||||| || || || |
Db     380 TCAGCATGACCACATGCTCCACACTGCTTGCCCTCGGAATGATGCCGCTGTGCCTCCTTA 439

Qy     440 TCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATTCCTTATCAGAACATAGGAA 499
      |||| ||| ||| || || || |||| ||| | ||||| ||
Db     440 TCTATACCAAATGTGGGTCGACTCTGGGAGCATCGTAATTCCTATGATAACATAGGTA 499
  
```

Qy 500 TTACCCCTTGTGTGCCTGACCATTCCCTGTGGCCTTTGGTGTCTATGTGAATTACAGATGGC 559
 | | | | | | | | | | | | | | | | | | | | | |
 Db 500 CATCTCTGGTTGCTCTCGTTGTTCCATTGGAATGTTGTTAATCACAAATGGC 559

Qy 560 CAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTTGGTGGGGTCCTCCTTCTGG 619
 | | | | | | | | | | | | | | | | | | | | | |
 Db 560 CCCAAAAGCAAAGATCATACTTAAATTTGGGTCCATCGCGGGCGCCATCCTCATTGTGC 619

Qy 620 TGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGATCTTGAATTCAGACATCACCCCTTC 679
 | | | | | | | | | | | | | | | | | | | | | |
 Db 620 TCATAGCTGTGGTTGGAGGAATATTGTACCAAAGCGCTGGATCATTGCTCCCAAATGT 679

Qy 680 TGACCATCAGTTTTCATCTTTCTTTGATTGGCCATGTCACGGGTTTTCTGCTGGCACTTT 739
 | | | | | | | | | | | | | | | | | | | | | |
 Db 680 GGATTATAGGAACAATATTTCTGTGGCGGGTTACTCCCTGGGGTTTTCTTCTGGCTAGAA 739

Qy 740 TTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCTTAGAAACTGGAGCTCAGAATA 799
 | | | | | | | | | | | | | | | | | | | | | |
 Db 740 TTGCTGGTCTACCCTGGTACAGGTGCCGAACGGTTGCTTTGAAACGGGGATGCAGAACA 799

Qy 800 TTCAGATGTGCATCACCATGCTCCAGTTATCTTTCCTGCTGAGCACTTGGTCCAGATGT 859
 | | | | | | | | | | | | | | | | | | | | | |
 Db 800 CGCAGCTATGTTCCACCATCGTTCAGCTCTCCTTCACTCCTGAGGAGCTCAATGTCTGAT 859

Qy 860 TGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGATGGATTTCCTTATTGTTGCAG 919
 | | | | | | | | | | | | | | | | | | | | | |
 Db 860 TCACCTTCCCGCTCATCTACAGCATTTTCCAGCTCGCCTTTGCCGCAATATTCTTAGGAT 919

Qy 920 CATATCAGACGTACAAGAGGAGATTGAAGAACAAACATGGAAAAA 964
 | | | | | | | | | | | | | | | | | | | | | |
 Db 920 TTTATGTGGCATACAAGAAATGTCATGGAAAAACAAGGCAGAAA 964

RESULT 8

CF998755

LOCUS CF998755 773 bp mRNA linear EST 25-NOV-2003
 DEFINITION AGENCOURT_16388570 NIH_ZGC_7 Danio rerio cDNA clone IMAGE:7040629
 5', mRNA sequence.

ACCESSION CF998755

VERSION CF998755.1 GI:38519606

KEYWORDS EST.

SOURCE Danio rerio (zebrafish)

ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.

REFERENCE 1 (bases 1 to 773)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Len Zon, Harvard

cDNA Library Preparation: Open Biosystems
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM14795 row: m column: 11
 High quality sequence stop: 663.

FEATURES	Location/Qualifiers
source	1. .773 /organism="Danio rerio" /mol_type="mRNA" /db_xref="taxon:7955" /clone="IMAGE:7040629" /tissue_type="whole body" /lab_host="DH10B" /clone_lib="NIH_ZGC_7" /note="Vector: pExpress1; Site_1: NotI; Site_2: EcoRV; Bulk tissue was collected from a whole adult individual from the Tuebingen strain. 1st strand cDNA was primed with a Not I - oligo(dT) primer, double-stranded cDNA was cloned into the Not I and EcoRV sites of pExpress-1. Library was size-selected for >1 kb fragments and normalized. A non-normalized version of this library is also available (NIH_ZGC_10). Library was constructed by Open Biosystems (Huntsville, AL)"

ORIGIN

Query Match 24.9%; Score 282.8; DB 14; Length 773;
 Best Local Similarity 61.7%; Pred. No. 1.2e-66;
 Matches 466; Conservative 0; Mismatches 288; Indels 1; Gaps 1;

Qy	115	ACTGTGATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTG	174
Db	10	ACCGTCATGTTGGCCATGTTTATGTTTTCAATGGGCTGCACTGTTGAGGCTAGAAACTG	69
Qy	175	TGGTCGCACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTC	234
Db	70	TGGGGGACGTTTCGCAGACCCTGGGGCATTTTTATAGGTTTCCTTTGCCAGTTTGGCATC	129
Qy	235	ATGCCTTTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATT	294
Db	130	ATGCCTTTCACAGCCTTCATACTTTCATTGCTTTTCAACGTGCTGCCAGTCCAGGCGGTG	189
Qy	295	GCTGTTCTCATCATGGGCTGCTGCCCCGGGGGCACCATCTCTAACATTTTCACCTTCTGG	354
Db	190	GTCATCATCATCATGGGCTGCTGCCCTGGAGGATCAAGCTCTAATGTTTTCTGCTACTGG	249
Qy	355	GTTGATGGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTG	414
Db	250	CTTGATGGAGACATGGACCTAAGCATCAGCATGACAGCGTGTTCTTCAATTTTGGCTCTG	309
Qy	415	GGAATGATGCCACTCTGCATTTATCTCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTC	474
Db	310	GGAATGATGCCTCTTTGTCTGCTCATTACACCACAATCTGGACTGCAGGCGATGCGATT	369
Qy	475	ACCATTCTTATCAGAACATAGGAATTACCCTTGTGTGCCTGACCATTCTGTGGCCTTT	534

```

      ||||| | | | | | | | | | | | | | | | | | |
Db    370 CAGATTCCTTACGACAATATTGGGATCACACTGGTGAGTTTGCTTGTGCCTGTCGGTCTT 429

Qy    535 GGTGTCTATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATTGGGGCC 594
      || | | | | | | | | | | | | | | | | | | | |
Db    430 GGGATGTTAGTGAAACACAAGTGGCCTAAAGCTGCCAAAAGATCCTCAAGGTTGGATCT 489

Qy    595 GTTGTGGTGGGGTCTCTCTTCTGGTGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGA 654
      || || | | | | | | | | | | | | | | | | | |
Db    490 GTGGTGGGAATTGTCTCATCATCGTCATTGCAGTAATTGGTGGTGTGCTTTATCAGTCC 549

Qy    655 TCTTGAATTTCAGACATCACCCCTTCTGACCATCAGTTTCATCTTTCCTTTGATTGGCCAT 714
      || |||| | | | | | | | | | | | | | | | | |
Db    550 TCATGGACCATTGCTCCCTCACTTTGGATCATTGGTACCATTATCCATTATTGGATT 609

Qy    715 GTCACGGGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGGACAATT 774
      | | | | | | | | | | | | | | | | | | | | |
Db    610 GGCTTAGGGTTCCTCTTGGCACGCTTTGTGGGCCAACCTTGGCACAGGTGCCGCACCATT 669

Qy    775 TCCTTAGAAACTGGAGCTCAGAATATTCAGATGTGCATCACCATGCTCCAGTTATCTTTC 834
      | | | | | | | | | | | | | | | | | | | | |
Db    670 GCTCTAGAAAC-GGCATGCAGAACGCCAGCTGGGCAGTACTATTTACCCAGTGTCTTT 728

Qy    835 ACTGCTGAGCACTTGGTCCAGATGTTGAGTTTCCC 869
      | | | | | | | | | | | | | | | | | |
Db    729 AGCCCTGCAGAGCTTGANGTCATGTTGCGGTTTCC 763

```

RESULT 9

BY779230

LOCUS BY779230 364 bp mRNA linear EST 10-DEC-2003

DEFINITION BY779230 RIKEN full-length enriched, 17.5 days embryo whole body
Mus musculus cDNA clone L930133F11 5', mRNA sequence.

ACCESSION BY779230

VERSION BY779230.1 GI:39705869

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 364)

AUTHORS Carninci,P., Waki,K., Shiraki,T., Konno,H., Shibata,K., Itoh,M.,
Aizawa,K., Arakawa,T., Ishii,Y., Sasaki,D., Bono,H., Kondo,S.,
Sugahara,Y., Saito,R., Osato,N., Fukuda,S., Sato,K., Watahiki,A.,
Hirozane-Kishikawa,T., Nakamura,M., Shibata,Y., Yasunishi,A.,
Kikuchi,N., Yoshiki,A., Kusakabe,M., Gustincich,S., Beisel,K.,
Pavan,W., Aidinis,V., Nakagawara,A., Held,W.A., Iwata,H., Kono,T.,
Nakauchi,H., Lyons,P., Wells,C., Hume,D.A., Fagiolini,M.,
Hensch,T.K., Brinkmeier,M., Camper,S., Hirota,J., Mombaerts,P.,
Muramatsu,M., Okazaki,Y., Kawai,J. and Hayashizaki,Y.

TITLE Targeting a complex transcriptome: the construction of the mouse
full-length cDNA encyclopedia

JOURNAL Genome Res. 13 (6B), 1273-1289 (2003)

MEDLINE 22703353

PUBMED 12819125

COMMENT Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site (<http://genome.gsc.riken.jp/>) for
further details.

FEATURES Location/Qualifiers
 source 1. .364
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="L930133F11"
 /tissue_type="whole body"
 /dev_stage="17.5 days embryo"
 /clone_lib="RIKEN full-length enriched, 17.5 days embryo
 whole body"

ORIGIN

Query Match 23.4%; Score 265.4; DB 13; Length 364;
Best Local Similarity 83.2%; Pred. No. 5.9e-62;
Matches 302; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

```

Qy      37  GCCAACAGTTCAGAGGAGGAGCTGCCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTC 96
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      2  GTCAACAGTACGGAGGAAGACCCGCGCTGGGAATGGAGGGCCATGCGAATCTAAAGCTG 61

Qy      97  GTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCC 156
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      62  CTTTTTACAGTGCTCTCGGCTGTGATGGTGGGTTTGGTCATGTTCTCTTTTGGATGTTCT 121

Qy     157  GTGGAGATCCGGAAGCTGTGGTCGCACATCAGGAGACCCTGGGGCATTGCTGTGGGACTG 216
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     122  GTGGAGAGTCAGAAGCTCTGTTGCACCTCAGAAGACCCTGGGGCATCGCAGTGGGCCTG 181

Qy     217  CTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTG 276
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     182  CTTTCCCAGTTTGGACTTATGCCTCTGACAGCTTATCTGTTAGCCATTGGCTTCGGTCTG 241

Qy     277  AAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCCCAGGGGGCACCATCTCT 336
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     242  AAACCATTCGAAGCTATTGCTGTCCTCATGATGGGGAGCTGCCCTGGGGGCACCATCTCT 301

Qy     337  AACATTTTTCACCTTCTGGGTTGATGGAGATATGGATCTCAGCATCAGTATGACAACCTGT 396
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     302  AATGTTCTCACCTTCTGGGTTGATGGAGATATGGATCTCAGCATCAGTATGACAACCTGT 361

Qy     397  TCC 399
      | | |

```

Db 362 TCC 364

RESULT 10

CB320835

LOCUS CB320835 834 bp mRNA linear EST 04-MAR-2003

DEFINITION AGENCOURT_12236884 NIH_MGC_136 Mus musculus cDNA clone

IMAGE:30289461 5', mRNA sequence.

ACCESSION CB320835

VERSION CB320835.1 GI:28845070

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 834)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Dr. David Rowe

cDNA Library Preparation: Invitrogen Corp

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: NDAM328 row: i column: 22

High quality sequence stop: 560.

FEATURES

source

Location/Qualifiers

1. .834

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="IMAGE:30289461"

/tissue_type="embryonic limb, maxilla and mandible"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_136"

/note="Vector: pCMV-SPORT6.1; Site_1: EcoRV; Site_2: NotI;

Normalized, full-length enriched library from pool of
mouse embryonic limb, maxilla and mandible, embryonic day

17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw
equivalents from respective days). Cloned directionally,

oligo-dT primed (5'-GACTAGTTCTAGATCGCGAGCGGCCGCC(T)15-3'.

Size selected for the >1kb fragments, average insert size

1.2 kb. Normalization to Cot 7.5 . Tissue contributed by

David Rowe; library constructed by ResGen, Invitrogen

Corp. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 22.6%; Score 256.4; DB 14; Length 834;

Best Local Similarity 72.5%; Pred. No. 2.5e-59;

Matches 364; Conservative 0; Mismatches 126; Indels 12; Gaps 2;

Qy

633 TGGTGTGGTCCTGGCGAAAGGATCTTGAATTCAGACATCACCCCTTCTGACCATCAGTTT 692

||| ||||||||| ||||| ||||| ||||| ||||| ||||| |||||

AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.

TITLE Direct Submission

JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA

COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment.

FEATURES Location/Qualifiers

source 1. .912

/organism="Pan troglodytes"

/mol_type="genomic DNA"

/db_xref="taxon:9598"

gene <1. .>912

/gene="SLC10A2"

/locus_tag="HCM5047"

ORIGIN

Query Match 21.6%; Score 245; DB 29; Length 912;

Best Local Similarity 47.4%; Pred. No. 3.6e-56;

Matches 386; Conservative 0; Mismatches 428; Indels 0; Gaps 0;

Qy 80 ATGGAACCTGGAGCTCGTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGT 139

|| | || | || | | || | || | || | || | || |

Db 80 ATAACATCCTAAGTGTGGTCCTAAGTACGGTGCTGACCATCCTGTTGGCCTTGGTGTGT 139

Qy 140 TCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTGCACATCAGGAGACCC'TGGG 199

|||| ||||| ||||| ||| | ||||| | | | | |||

Db 140 TCTCCATGGGATGCAACGTGGAAATCAAGAAATTTCTAGGGCACATAAAGCGCCGTGGG 199

Qy 200 GCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGG 259

|||| ||| || | |||| ||||| ||||| | ||| | ||||

Db 200 GCATTTGTGTTGGCTTCCTCTGTCAGTTTGAATCATGCCCCCTCACAGGATTCATCCTGT 259

Qy 260 CCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCC 319

| | ||| | || |||| || | | || |||| | | | |||||

Db 260 CGGTGGCCTTTGACATCCTCCCGCTCCAGGCCGTGGTGGTGTCTATTATAGGATGCTGCC 319

Qy 320 CGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGATGGAGATATGGATCTCAGCA 379

| || || | || || || || || || | || |

Db 320 CTGGAGGAAC'TGCCCTCAATATCTTGGCCTATTNGGTCGNNNNNNNNNNNNNNNNNNCG 379

Qy 380 TCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATC 439

|||| |||| || || |||| || |||| ||||| || ||| | |

Db 380 TCAGCATGACCACATGCTCCACACTGCTTGCCCTCGGAATGATGCCGCTGTGCCTCCTTA 439

Qy 440 TCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATTCCCTTATCAGAACATAGGAA 499

|||| ||| || || || || || |||| || | |||||

Db 440 TCTATACCAAATGTGGGTCGACTCTGGGAGCATCGTAATCCCTATGATAACATAGNNN 499

Qy 500 TTACCCTTGTGTGCCTGACCATTCCCTGTGGCCTTTGGTGTCTATGTGAATTACAGATGGC 559

' || || || | |||| || |||| | | || ||| ||| ||||

Db 500 NNNNNCTGGTTGCTCTC'NNNTNCCTGTTTCCATTGGGATGTTTGTTAATCACAAATGGC 559

Qy 560 CAAACAATCCAAATCATTCTCAAGATTGGGGCCGTGTTGGTGGGGTCCTCCTTCTGG 619

| || || | | |||| || | | || | |||| |

Db	560	CCCCAAAAGCAANNNNNNNNNNNNNNNATTGGGTCCATCGCGGGGCCCATCCTCATTNNNN	619
Qy	620	TGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGATCTTGGAATTCAGACATCACCCCTTC	679
Db	620	NN	679
Qy	680	TGACCATCAGTTTCATCTTTCCTTTGATTGGCCATGTCACGGGTTTTCTGCTGGCACTTT	739
Db	680	NN	739
Qy	740	TTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCCCTTAGAAACTGGAGCTCAGAATA	799
Db	740	NNNNNNNNNNNNNNNNNNNNNNNGTGCCGAACGGTTGCT'TTTGAAACGGGGATGCAGAACA	799
Qy	800	TTCAGATGTGCATCACCATGCTCCAGTTATCTTTCACTGCTGAGCACTTGGTCCAGATGT	859
Db	800	CGCAGCTATGTTCCACCATCGTTCAGCTGTCCTTCACTCCTGAGGAGCTCAATGTCTGAT	859
Qy	860	TGAGTTTCCCCTGAGCTATGGACTCTTCCAGCT	893
Db	860	TCACCTTCCCGCTCATCTACAGCATTTTCCAGCT	893

Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

MEDLINE 22354683

PUBMED 12466851

COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES Location/Qualifiers

source 1. .372
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="L930044F15"
/tissue_type="whole body"
/dev_stage="17.5 days embryo"
/clone_lib="RIKEN full-length enriched, 17.5 days embryo whole body"

ORIGIN

Query Match 21.2%; Score 240.6; DB 13; Length 372;
 Best Local Similarity 79.4%; Pred. No. 4.2e-55;
 Matches 285; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

```

Qy      1 ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG 60
      ||||| | | ||| | ||| ||| ||||| ||||| ||||| ||| |
Db      14 ATGAGCACAGACTGTGCGGGCAACTCCACCTGCCCTGTCAACAGTACGGAGGAAGACCCG 73

Qy     61 CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTG 120
      || ||||| ||||| |||| | || | |||| | |||| | || |||||
Db     74 CCCGTGGGAATGGAGGGCCATGCGAATCTAAAGCTGCTTTTTACAGTGCTCTCGGCTGTG 133

Qy    121 ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTGC 180
      ||| |||| | ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    134 ATGGTGGGTTTGGTCATGTTCTCTTTGGATGTTCTGTGGAGAGTCAGAAGCTCTGGTTG 193

Qy    181 CACATCAGGAGACCCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT 240
      ||| ||| ||||| ||||| | |||| | |||| | ||||| || |||||
Db    194 CACCTCACAAGACCCCTGGGGCATCCAGTGGGCCTGCTTTCCAGTTTGGACTTATGCCT 253

Qy    241 TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT 300
      | ||||| ||||| | ||||| |||| | |||| | ||||| ||||| |||||
Db    254 CTGACAGCTTATCTGTTAGCCATTGGCTTCGGTCTGAAACCATTCCAAGCTATTGCTGTC 313

Qy    301 CTCATCATGGGCTGCTGCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGA 359
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    314 CTCATGATGGGAGCTGCCCT'GGGGGCACCATCTCTAATGTTCTCACCTTCTGGGTTGA 372
  
```

RESULT 13
 CA353647

LOCUS CA353647 620 bp mRNA linear EST 05-NOV-2002
 DEFINITION 625196 NCCCWA 1RT Oncorhynchus mykiss cDNA clone 1RT74B18_D_A09 5',
 mRNA sequence.
 ACCESSION CA353647
 VERSION CA353647.1 GI:24598818
 KEYWORDS EST.
 SOURCE Oncorhynchus mykiss (rainbow trout)
 ORGANISM Oncorhynchus mykiss
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
 REFERENCE 1 (bases 1 to 620)
 AUTHORS Rexroad,C.E. and Keele,J.W.
 TITLE Sequence analysis of a rainbow trout normalized cDNA library
 JOURNAL Unpublished (2002)
 COMMENT Contact: Rexroad CE
 USDA, ARS, National Center for Cool and Cold Water Aquaculture
 11876 Leetown Road, Kearneysville, WV 25430, USA
 Tel: 304 724 8340 x2129
 Fax: 304 725 0351
 Email: crexroad@ncccwa.ars.usda.gov
 Single pass sequencing. Bases called with phred v0.020425.c and
 trimmed with the aid of the trim_alt option. Vector identified by
 cross_match v0.990329.
 Seq primer: AGCGGATAACAATTTACACAGGA.
 FEATURES Location/Qualifiers

```

source      1. .620
            /organism="Oncorhynchus mykiss"
            /mol_type="mRNA"
            /db_xref="taxon:8022"
            /clone="1RT74B18_D_A09"
            /tissue_type="pooled"
            /lab_host="DH10B"
            /clone_lib="NCCCWA 1RT"
            /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
            Library made from pooled tissue from brain, gill, liver,
            spleen, muscle, and kidney."

```

ORIGIN

```

Query Match      20.1%; Score 227.6; DB 14; Length 620;
Best Local Similarity 62.6%; Pred. No. 2e-51;
Matches 387; Conservative 0; Mismatches 229; Indels 2; Gaps 2;

```

```

Qy      88 CTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGTTCTCTTTG 147
      |||  || ||| |||  | |||  || || ||| |||  || ||||| |||  ||
Db      1 CTGAGCCTAGTTCTCAGCATCGTGCTGACCGTCATGCTGGCCATGGTCATGTTCTCCATG 60

Qy     148 GGATGTTCCGTGGAGATCCGGAAGCTGTGGTTCGCACATCAGGAGACCCTGGGGCATTGCT 207
      || ||  ||||| |||  | ||||| |||  ||||| ||| || ||||| |||  |
Db      61 GGCTGCACCGTGGAGGCCGGAAGCTGTGGGGACACATCAAGAGGCCATGGGGAATTTTT 120

Qy     208 GTGGGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGGCCATTAGC 267
      | ||  | | ||||| |||  | | | | | ||| |||  |||| | | | |
Db     121 ATCGGCTTCTTGTGCCAGTTCGGCATTG-GTCCTTCACCGCCTTCGCCCTGTCTGCTGGCC 179

Qy     268 TTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCCCCGGGGGGC 327
      ||  ||  ||| || || ||  | |  | ||||| ||||| ||| || |||
Db     180 TTCAACGTGCTGCCCCTGCAGGCCGTCGTCATCATCATCATGGGCTGCTGTCCCGGTGGC 239

Qy     328 ACCATCTCTAACATTTTTCACCTTCTGGGTTGATGGAGATATGGAT-CTCAGCATCAGTAT 386
      ||| ||||| ||  | ||| |||| | ||||| |||| | |||| | |||| ||
Db     240 TCCAGCTCTAATATCATTGCCTACTGGCTGGATGGAGACATGGCTCCTCAGTATCAGCAT 299

Qy     387 GACAACCTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATCTCTACAC 446
      |||| |||| ||| | |  ||||| ||||| || || |  |||||
Db     300 GACAGCCTGCTCCTCTATCCTGGCCCTGGGGATGATGCCTCTGTGTCTGCTCATCTACAC 359

Qy     447 CTGGTCCTGGAGTCTTCAGCAGAATCTCACCATTTCCTTATCAGAACATAGGAATTACCCT 506
      |  ||||  |  | |  ||  || || || || || || || || || ||
Db     360 GTCTGTCTGGACCTCTGCTGACACCATCCAGATCCCCCTACCAAAGCATAGGTATCACTTT 419

Qy     507 TGTGTGCCTGACCATTCTGTGGCCTTTGGTGTCTATGTGAATTACAGATGGCCAAAACA 566
      |||| |||  ||| |||| ||| | ||  ||| || || ||| |||| ||
Db     420 GGTGTCCCTCCTCATCCCTGTGCCCCTGGGAATCTACGTCAAAAACAAGTGGCCCGAAAT 479

Qy     567 ATCCAAAATCATTCTCAAGATTGGGGCCGTTGTTGGTGGGGTCTCCTTCTGGTGGTTCGC 626
      | | |||| || ||||| | || || | ||||  |||| | | | | ||
Db     480 AGCTAAAAAGATCCTCAAGGTGGGTTCCATAGTTGGCCTCCTCCTCATCATATAATAGC 539

Qy     627 AGTTGCTGGTGTGGTCTCTGGCGAAAGGATCTTGAATTCAGACATCACCCCTTCTGACCAT 686
      || | |||| ||| |||  |  | ||||  || || || || ||
Db     540 GGTGGTGGTGGGGTGTGTACCAGTCCTTCTGGACCATCTCTCCCTCTCTCTGGATCAT 599

```


Qy 687 CAGTTTCATCTTTCCTTT 704
 | | | | | | |
 Db 600 CGGAGCCATCTACCCCTT 617

RESULT 14

BY135403

LOCUS BY135403 378 bp mRNA linear EST 09-DEC-2002

DEFINITION BY135403 RIKEN full-length enriched, 17.5 days embryo whole body
 Mus musculus cDNA clone L930061J09 5', mRNA sequence.

ACCESSION BY135403

VERSION BY135403.1 GI:26270955

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 378)

AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
 Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
 Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
 Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
 Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
 Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
 Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
 Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
 Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
 Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
 Kawaji,H., Kawasaki,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
 Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
 Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
 Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
 Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
 Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
 Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
 Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
 Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
 Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
 Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
 Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
 Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
 Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
 Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
 Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
 Rogers,J., Birney,E. and Hayashizaki,Y.

TITLE Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

MEDLINE 22354683

PUBMED 12466851

COMMENT Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,
 URL: <http://genome.gsc.riken.go.jp/>
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
 Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
 Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
 Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
 Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
 Hayashizaki, Y. Direct Submission
 Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for
 further details.

FEATURES Location/Qualifiers
 source 1. .378
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="L930061J09"
 /tissue_type="whole body"
 /dev_stage="17.5 days embryo"
 /clone_lib="RIKEN full-length enriched, 17.5 days embryo
 whole body"

ORIGIN

Query Match 20.0%; Score 227.2; DB 13; Length 378;
 Best Local Similarity 78.8%; Pred. No. 2.1e-51;
 Matches 271; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```

Qy      1 ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG 60
      ||||| | | ||| | ||| ||| ||||| ||||| ||||| ||| |
Db      35 ATGAGCACAGACTGTGCGGGCAACTCCACCTGCCCTGTCAACAGTACGGAGGAAGACCCG 94

Qy      61 CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTG 120
      || ||||| ||||| |||| || || || |||| |||| ||||| ||| |||||
Db      95 CCCGTGGGAATGGAGGGCCATGCGAATCTAAAGCTGCTTTTTACAGTGCTCTCGGCTGTG 154

Qy     121 ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTTCG 180
      ||| |||| || ||||| ||||| ||||| ||||| ||||| ||||| |||
Db     155 ATGGTGGGTTTGGTCATGTTCTCTTTTGGATGTTCTGTGGAGAGTCAGAAGCTCTGGTTG 214

Qy     181 CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT 240
      ||| |||| ||||| ||||| ||| | ||| ||||| | ||||| ||| |||||

```


Matches 341; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

```
Qy      95 TCGTTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGTTCTCTTTGGGATGTT 154
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     264 TTGTGATGAGCGTTGCCATTACCGTCATGTTGGCCATGTTTATGTTTTCAATGGGCTGCA 323

Qy     155 CCGTGGAGATCCGGAAGCTGTGGTCGCACATCAGGAGACCCTGGGGCATTGCTGTGGGAC 214
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     324 CTGTTGAGGCTAGAAAACGTGTGGGGGCACGTTTCGCAGACCCTGGGGCATTTTTATAGGTT 383

Qy     215 TGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGGCCATTAGCTTTTCTC 274
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     384 TCCTTTGCCAGTTTGGCATCATGCCTTTCACAGCCTTCATACTTTCATTGCTTTTCAACG 443

Qy     275 TGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCCCGGGGGGCACCATCT 334
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     444 TGCTGCCAGTCCAGGCGGTGGTCATCATCATCATGGGCTGCTGCCCTGGAGGATCAAGCT 503

Qy     335 CTAACATTTTACCTTCTGGGTTGATGGAGATATGGATCTCAGCATCAGTATGACAACCT 394
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     504 CTAATGTTTTCTGCTACTGGCTTGATGGAGACATGGACCTAAGCATCAGCATGACAGCGT 563

Qy     395 GTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATCTCTACACCTGGTCTT 454
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     564 GTTCTTCAATTTTGGCTCTGGGAATGATGCCTCTTTGTCTGCTCATCTACACCACAATCT 623

Qy     455 GGAGTCTTCAGCAGAATCTCACCATTCTTATCAGAACATAGGAATTACCCTTGTGTGCC 514
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     624 GGACTGCAGGCGATGCGATCCAGATTCTTACGACAATATTGGGATCACACTGGTGAGTT 683

Qy     515 TGACCATTCTGTGGCCTTTGGTGTCTATGTGAATTACAGATGGCCAAAACAATCCAAAA 574
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     684 TGCTTGTGCCTGTCGGTCTTGGGATGTTAGTGAAACACAAGTGGCCTAAAGCTGCCAAAA 743

Qy     575 TCATTCTCAAGATTGGGGCCGTTGTTGGTGGGGTCCTCCTTCTGGTGGTTCGAGT 629
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     744 AGATCCTNCAGGTTGGATCTGTGGTGGGAATCGTCCTCATCATCGTCATTGAGT 798
```

Search completed: March 25, 2004, 18:53:51

Job time : 3450 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2004, 12:53:57 ; Search time 4659 Seconds
(without alignments)
10549.689 Million cell updates/sec

Title: US-10-091-628-1
Perfect score: 1134
Sequence: 1 atgagagccaattgttccag.....acatcacttcatgtgaatag 1134

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*

28: em_un:*
 29: em_vi:*
 30: em_htg_hum:*
 31: em_htg_inv:*
 32: em_htg_other:*
 33: em_htg_mus:*
 34: em_htg_pln:*
 35: em_htg_rod:*
 36: em_htg_mam:*
 37: em_htg_vrt:*
 38: em_sy:*
 39: em_htgo_hum:*
 40: em_htgo_mus:*
 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result			%		Query		DB	ID	Description
	No.	Score	Match	Length	Match	Length			
	1	1134	100.0	1134	9	AJ583502			AJ583502 Homo sapi
	2	1130.8	99.7	1517	6	AX575470			AX575470 Sequence
	3	704.4	62.1	1122	10	AJ583504			AJ583504 Mus muscu
	4	657.2	58.0	1113	10	AJ583503			AJ583503 Rattus no
	5	655.8	57.8	987	6	AX574600			AX574600 Sequence
c	6	377	33.2	23618	9	AC079237			AC079237 Homo sapi
c	7	377	33.2	192263	9	AC093827			AC093827 Homo sapi
	8	375.4	33.1	65268	2	AC099847			AC099847 Homo sapi
	9	320.4	28.3	2263	6	AR033870			AR033870 Sequence
	10	320.4	28.3	2263	6	I32744			I32744 Sequence 1
	11	320.4	28.3	2263	10	CGU02028			U02028 Cricetulus
	12	309.6	27.3	1916	5	BC053189			BC053189 Danio rer
	13	307.8	27.1	974	10	D87059			D87059 House mouse
	14	307.8	27.1	1629	10	AB002693			AB002693 Mus muscu
	15	306.4	27.0	1116	4	OCSDBATRP			Z54357 O.cuniculus
	16	298.4	26.3	1047	4	CFA581082			AJ581082 Canis fam
	17	297.8	26.3	1047	6	AR033871			AR033871 Sequence
	18	297.8	26.3	1047	6	I32745			I32745 Sequence 3
	19	297.8	26.3	3779	6	AX589492			AX589492 Sequence
	20	297.8	26.3	3779	9	HSU10417			U10417 Homo sapien
	21	293.2	25.9	4269	10	RNU07183			U07183 Rattus norv
	22	265	23.4	243333	2	AC120684			AC120684 Rattus no
	23	265	23.4	247127	2	AC098523			AC098523 Rattus no
	24	261.8	23.1	215210	10	AL713989			AL713989 Mouse DNA
c	25	217	19.1	65958	9	AC105413			AC105413 Homo sapi
	26	184	16.2	1212	4	OCU131361			AJ131361 Oryctolag
	27	183.2	16.2	1663	6	AX401950			AX401950 Sequence
	28	183.2	16.2	1663	6	AX827529			AX827529 Sequence
	29	183.2	16.2	1663	10	RATSBACT			M77479 Rattus norv
c	30	182.6	16.1	543	11	G51602			G51602 SHGC-79180
	31	173.8	15.3	1411	10	MMU95132			U95132 Mus musculu
	32	173.8	15.3	1579	10	MMU95131			U95131 Mus musculu
	33	173.8	15.3	1596	10	AB003303			AB003303 Mouse mRN

34	173.8	15.3	1649	10	BC021154	BC021154 Mus muscu
35	173.6	15.3	1580	6	AX409529	AX409529 Sequence
36	173.6	15.3	1580	9	HUMNTCP	L21893 Human Na/ta
37	141.4	12.5	1988	6	AX921120	AX921120 Sequence
38	138.2	12.2	1437	9	AK126542	AK126542 Homo sapi
39	138.2	12.2	1686	9	BC012048	BC012048 Homo sapi
40	138.2	12.2	1705	9	BC019066	BC019066 Homo sapi
41	133.2	11.7	976	4	AY292653	AY292653 Oryctolag
42	133.2	11.7	27247	4	OCAJ2005	AJ002005 Oryctolag
43	131	11.6	152080	5	AL953877	AL953877 Zebrafish
44	118.8	10.5	2020	6	AX589493	AX589493 Sequence
45	118.8	10.5	2020	9	HSISDBA1	U67669 Human ileal

ALIGNMENTS

RESULT 1

AJ583502

LOCUS AJ583502 1134 bp mRNA linear PRI 24-SEP-2003

DEFINITION Homo sapiens mRNA for sodium-dependent organic anion transporter (SOAT gene).

ACCESSION AJ583502

VERSION AJ583502.1 GI:35208820

KEYWORDS SOAT gene; sodium-dependent organic anion transporter.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Geyer, J. and Petzinger, E.

TITLE cloning of a sodium-dependent organic anion transporter (SOAT) from human adrenal gland

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1134)

AUTHORS Geyer, J.

TITLE Direct Submission

JOURNAL Submitted (23-SEP-2003) Geyer J., Institute of Pharmacology and Toxicology, University of Giessen, Frankfurter Str. 107, 35392 Giessen, GERMANY

FEATURES Location/Qualifiers

source 1. .1134

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/chromosome="4"

/tissue_type="adrenal gland"

gene 1. .1134

/gene="SOAT"

CDS 1. .1134

/gene="SOAT"

/codon_start=1

/evidence=experimental

/product="sodium-dependent organic anion transporter"

/protein_id="CAE47477.1"

/db_xref="GI:35208821"

/translation="MRANCSSSSACPANSSEEEELPVGLEVHGNLELVFTVVSTVMMGL"

LMFSLGCSVEIRKLWSHIRRPWGIAVGLLCQFGLMPFTAYLLAISFSLKPVQAI AVLI
MGCCPGGTISNIFTFWVDGMDLSISMTTCSTVAALGMMPLCIYLYTWSWSLQQNLTI
PYQNI GITLVCLTIPVAFGVVYNYRWPQSKIILKIGAVVGGVLLL VVAVAGVV LAKG
SWNSDITLLTISFIFPLIGHVTGFL LALFTHQSWQRCRTISLETGAQNIQMCITMLQL
SFTA EHLVQMLS FPLAYGLFQLIDGFLIVAAYQTYKRRLKNKHGKKN SGCTEVCHTRK
STSSRETNAFLEVNEEGAITPGPPGMDCHRALEPVGHITSCE"

ORIGIN

Query Match 100.0%; Score 1134; DB 9; Length 1134;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG	60
Db	1	ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG	60
Qy	61	CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTG	120
Db	61	CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTG	120
Qy	121	ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCG	180
Db	121	ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCG	180
Qy	181	CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT	240
Db	181	CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT	240
Qy	241	TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT	300
Db	241	TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT	300
Qy	301	CTCATCATGGGCTGCTGCCCGGGGGGACCATCTCTAACATTTTCACCTTCTGGGTGAT	360
Db	301	CTCATCATGGGCTGCTGCCCGGGGGGACCATCTCTAACATTTTCACCTTCTGGGTGAT	360
Qy	361	GGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATG	420
Db	361	GGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATG	420
Qy	421	ATGCCACTCTGCATTTATCTCTACACCTGGTCTCGGAGTCTTCAGCAGAATCTCACCATT	480
Db	421	ATGCCACTCTGCATTTATCTCTACACCTGGTCTCGGAGTCTTCAGCAGAATCTCACCATT	480
Qy	481	CCTTATCAGAACATAGGAATTACCCTTGTGTGCCTGACCATTCTGTGGCCTTTGGGTGTC	540
Db	481	CCTTATCAGAACATAGGAATTACCCTTGTGTGCCTGACCATTCTGTGGCCTTTGGGTGTC	540
Qy	541	TATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTT	600
Db	541	TATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTT	600
Qy	601	GGTGGGGTCTCTCTTCTGGTGGTCGAGTTGCTGGTGTGGTCTGGCGAAAGGATCTTGG	660
Db	601	GGTGGGGTCTCTCTTCTGGTGGTCGAGTTGCTGGTGTGGTCTGGCGAAAGGATCTTGG	660
Qy	661	AATTCAGACATCACCTTCTGACCATCAGTTTCATCTTTCCTTTGATTGGCCATGTCACG	720


```

      |||
Db      661 AATTCAGACATCACCTTCTGACCATCAGTTTCATCTTTCCTTTGATTGGCCATGTCACG 720
Qy      721 GGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCCCTTA 780
      |||
Db      721 GGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCCCTTA 780
Qy      781 GAAACTGGAGCTCAGAATATTCAGATGTGCATCACCATGCTCCAGTTATCTTTCACTGCT 840
      |||
Db      781 GAAACTGGAGCTCAGAATATTCAGATGTGCATCACCATGCTCCAGTTATCTTTCACTGCT 840
Qy      841 GAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGAT 900
      |||
Db      841 GAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGAT 900
Qy      901 GGATTTCCTTATTGTTGCAGCATATCAGACGTACAAGAGGAGATTGAAGAACAAACATGGA 960
      |||
Db      901 GGATTTCCTTATTGTTGCAGCATATCAGACGTACAAGAGGAGATTGAAGAACAAACATGGA 960
Qy      961 AAAAAGAACTCAGGTTGCACAGAAGTCTGCCATACGAGGAAATCGACTTCTTCCAGAGAG 1020
      |||
Db      961 AAAAAGAACTCAGGTTGCACAGAAGTCTGCCATACGAGGAAATCGACTTCTTCCAGAGAG 1020
Qy      1021 ACCAATGCCTTCTTGGAGGTGAATGAAGAAGGTGCCATCACTCCTGGGCCACCAGGGCCA 1080
      |||
Db      1021 ACCAATGCCTTCTTGGAGGTGAATGAAGAAGGTGCCATCACTCCTGGGCCACCAGGGCCA 1080
Qy      1081 ATGGATTGCCACAGGGCTCTCGAGCCAGTTGGCCACATCACTTCATGTGAATAG 1134
      |||
Db      1081 ATGGATTGCCACAGGGCTCTCGAGCCAGTTGGCCACATCACTTCATGTGAATAG 1134

```

RESULT 2

AX575470

LOCUS AX575470 1517 bp DNA linear PAT 07-JAN-2003

DEFINITION Sequence 26 from Patent WO02077237.

ACCESSION AX575470

VERSION AX575470.1 GI:27552072

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Lee,E.A., Ding,L., Baughn,M.R., Tribouley,C.M., Bruns,C.M.,
Elliott,V.S., Walia,N.K., Forsythe,I.J., Raumann,B.E., Burford,N.,
Lal,P.G., Thornton,M., Gandhi,A.R., Arvizu,C., Yao,M.G., Yue,H.,
Xu,Y., Hafalia,A.J. and Ison,C.H.

TITLE Transporters and ion channels

JOURNAL Patent: WO 02077237-A 26 03-OCT-2002;

Incyte Genomics, Inc. (US)

FEATURES Location/Qualifiers

source 1. .1517

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

/note="Incyte ID No: 7472881CB1"

ORIGIN

Query Match 99.7%; Score 1130.8; DB 6; Length 1517;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1132; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	1	ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG	60
Db	249	ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG	308
Qy	61	CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTACAGTGGTGTCCACTGTG	120
Db	309	CCAGTGGGACTGGAGGCGCATGGAAACCTGGAGCTCGTTTTACAGTGGTGTCCACTGTG	368
Qy	121	ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCG	180
Db	369	ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCG	428
Qy	181	CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTGGGGCTCATGCCT	240
Db	429	CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTGGGGCTCATGCCT	488
Qy	241	TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT	300
Db	489	TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT	548
Qy	301	CTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGAT	360
Db	549	CTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGAT	608
Qy	361	GGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATG	420
Db	609	GGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATG	668
Qy	421	ATGCCACTCTGCATTTATCTCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATT	480
Db	669	ATGCCACTCTGCATTTATCTCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATT	728
Qy	481	CCTTATCAGAACATAGGAATTACCCTTGTGTGCCTGACCATTCTGTGGCCTTTGGTGTC	540
Db	729	CCTTATCAGAACATAGGAATTACCCTTGTGTGCCTGACCATTCTGTGGCCTTTGGTGTC	788
Qy	541	TATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTT	600
Db	789	TATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTT	848
Qy	601	GGTGGGGTCCTCCTTCTGGTGGTGCAGTTGCTGGTGTGGTCTGGCGAAAGGATCTTGG	660
Db	849	GGTGGGGTCCTCCTTCTGGTGGTGCAGTTGCTGGTGTGGTCTGGCGAAAGGATCTTGG	908
Qy	661	AATTCAGACATCACCCCTTCTGACCATCAGTTTCATCTTTCCTTTGATTGGCCATGTCACG	720
Db	909	AATTCAGACATCACCCCTTCTGACCATCAGTTTCATCTTTCCTTTGATTGGCCATGTCACG	968
Qy	721	GGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTCCTTA	780
Db	969	GGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTCCTTA	1028

```

Qy      781 GAAACTGGAGCTCAGAATATTCAGATGTGCATCACCATGCTCCAGTTATCTTTCACTGCT 840
        |||
Db      1029 GAAACTGGAGCTCAGAATATTCAGATGTGCATCACCATGCTCCAGTTATCTTTCACTGCT 1088

Qy      841 GAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGAT 900
        |||
Db      1089 GAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGAT 1148

Qy      901 GGATTTCTTATTGTTGCAGCATATCAGACGTACAAGAGGAGATTGAAGAACAAACATGGA 960
        |||
Db      1149 GGATTTCTTATTGTTGCAGCATATCAGACGTACAAGAGGAGATTGAAGAACAAACATGGA 1208

Qy      961 AAAAAGAACTCAGGTTGCACAGAAGTCTGCCATACGAGGAAATCGACTTCTTCCAGAGAG 1020
        |||
Db      1209 AAAAAGAACTCAGGTTGCACAGAAGTCTGCCATACGAGGAAATCGACTTCTTCCAGAGAG 1268

Qy      1021 ACCAATGCCTTCTTGGAGGTGAATGAAGAAGGTGCCATCACTCCTGGGCCACCAGGGCCA 1080
        |||
Db      1269 ACCAATGCCTTCTTGGAGGTGAATGAAGAAGGTGCCATCACTCCTGGGCCACCAGGGCCA 1328

Qy      1081 ATGGATTGCCACAGGGCTCTCGAGCCAGTTGGCCACATCACTTCATGTGAATAG 1134
        |||
Db      1329 ATGGATTGCCACAGGGCTCTCGAGCCAGTTGGCCACATCACTTCATGTGAATAG 1382

```

RESULT 3

AJ583504

LOCUS AJ583504 1122 bp mRNA linear ROD 24-SEP-2003

DEFINITION Mus musculus mRNA for sodium-dependent organic anion transporter (SOAT gene).

ACCESSION AJ583504

VERSION AJ583504.1 GI:35208824

KEYWORDS SOAT gene; sodium-dependent organic anion transporter.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1

AUTHORS Geyer, J., Godoy, J.R. and Petzinger, E.

TITLE Cloning of a sodium-dependent organic anion transporter (SOAT) from mouse liver

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1122)

AUTHORS Geyer, J.

TITLE Direct Submission

JOURNAL Submitted (23-SEP-2003) Geyer J., Institute of Pharmacology and Toxicology, University of Giessen, Frankfurter Str. 107, 35392 Giessen, GERMANY

FEATURES Location/Qualifiers

source 1. .1122
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /chromosome="5"
 /tissue_type="liver"

```

gene      1. .1122
          /gene="SOAT"
CDS       1. .1122
          /gene="SOAT"
          /codon_start=1
          /product="sodium-dependent organic anion transporter"
          /protein_id="CAE47479.1"
          /db_xref="GI:35208825"
          /translation="MSTDCAGNSTCPVNSTEEDPPVGMEGHANLKLFTVLSAVMVGL
VMFSFGCSVESQKLWLHLRRPWGIAVGLLSQFGLMPLTAYLLAIGFGLKPFQAIIVLM
MGSCPGGTISNVLTFWVDGMDLSISM TTCSTVAALGMMPLCLYIYTRSWTLTQNLVI
PYQSIGITLVSLVVPVASGVYVNYRWPQATVILKVGAILGGMLLVAVTGMVLAKG
WNTDVTLLVISCIFPLVGHVTGFLLAFLTHQSWQRCRTIS IETGAQNIQLCIAMLQLS
FSAEYLVQLLNALAYGLFQVLHGLLIVAAYQAYKRRQKSKCRRQHPDCPDVCYEQP
RETS AFLDKGDEAAVT LGPVQPEQHHRAAELTSHIP SCE"

```

ORIGIN

```

Query Match      62.1%;  Score 704.4;  DB 10;  Length 1122;
Best Local Similarity  77.7%;  Pred. No. 2.2e-204;
Matches 881;  Conservative 0;  Mismatches 241;  Indels 12;  Gaps 2;

```

```

Qy      1 ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG 60
        ||||| | | ||| | ||| ||| ||||| ||||| | ||||| || | |
Db      1 ATGAGCACAGACTGTGCGGGCAACTCCACCTGCCCTGTCAACAGTACGGAGGAAGACCCG 60

Qy     61 CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTTACAGTGGTGTCCACTGTG 120
        || ||||| ||||| |||| | | ||| |||| ||||| | |||||
Db     61 CCCGTGGGAATGGAGGGCCATGCGAATCTAAAGCTGCTTTTTACAGTGCTCTCGGCTGTG 120

Qy    121 ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTTCG 180
        ||| |||| | | ||||| ||||| ||||| ||||| | ||||| |||| |
Db    121 ATGGTGGGTTTGGTCATGTTCTCTTTGGATGTTCTGTGGAGAGTCAGAAGCTCTGGTTG 180

Qy    181 CACATCAGGAGACCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT 240
        ||| |||| ||||| ||||| || | ||||| |||| | ||||| || |||||
Db    181 CACCTCAGAAGACCTGGGGCATCGCAGTGGGCCTGCTTTCCAGTTTGGACTTATGCCT 240

Qy    241 TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT 300
        | ||||| |||| | ||||| |||| | ||||| ||| ||||| ||||| |||||
Db    241 CTGACAGCTTATCTGTTAGCCATTGGCTTCGGTCTGAAACCATTCCAAGCTATTGCTGTC 300

Qy    301 CTCATCATGGGCTGCTGCCCGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGAT 360
        ||||| |||| | ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    301 CTCATGATGGGAGCTGCCCTGGGGGCACCATCTCTAATGTTCTCACCTTCTGGGTTGAT 360

Qy    361 GGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATG 420
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    361 GGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACAGTGGCCGCCCTGGGAATG 420

Qy    421 ATGCCACTCTGCATTTATCTCTACACCTGGTCCCTGGAGTCTTCAGCAGAATCTCACCATT 480
        ||||| ||||| | || | ||||| ||||| ||||| ||| ||||| ||| |||||
Db    421 ATGCCTCTCTGCCTCTACATCTACACCGGTCTGGACTCTGACACAGAACCTCGTCATT 480

Qy    481 CCTTATCAGAACATAGGAATTACCTTGTGTGCCTGACCATTCTGTGGCCTTTGGTGTC 540
        || ||||| ||||| ||||| ||||| ||||| ||||| ||| ||| |||
Db    481 CCGTATCAGAGCATAGGAATTACCTTGTGTCCCTGGTGGTTCTGTGGCTTCTGGCGTC 540

```


Qy 301 CTCATCATGGGCTGCTGCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGAT 360
 ||||| ||||| ||||| ||||| | ||||| ||||| |||||
 Db 301 CTCATCATGGGAGCTGCCCTGGGGGCACCGTCTCTAATGTCTCACCTTCTGGGTTGAT 360

Qy 361 GGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATG 420
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 361 GGAGATATGGACCTCAGCATCAGCATGACGACCTGCTCCACAGTGGCTGCTCTGGGAATG 420

Qy 421 ATGCCACTCTGCATTTATCTCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATT 480
 ||||| ||||| | || ||||| ||||| ||||| ||||| |||||
 Db 421 ATGCCCTCTGCCTCTACGTCTACACCGGTCTGGACTCTTCACAGAGCCTCACCATC 480

Qy 481 CCTTATCAGAACATAGGAATTACCCTTGTGTGCCTGACCATTCTGTGGCCTTTGGTGTC 540
 || || |||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 481 CCGTACCAGAGCATAGGAATTACCCTTGTGTCCCTGGTTGTTCTGTGGCCTCCGGCATC 540

Qy 541 TATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTT 600
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 541 TATGTGAATTATAGGTGGCCAAAGCAAGCAACATTCATTCTCAAGGTGGGGCTGCTGTT 600

Qy 601 GGTGGGGTCCTCCTTCTGGTGGTGCAGTTGCTGGTGTGGTCCTGGCGAAAGGATCTTGG 660
 || || | |||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 601 GGCGGCATGCTCCTCCTGGTGGTGGCAGTTACCGGCGTGGTCCTGGCAAAGGG---CTGG 657

Qy 661 AATTCAGACATCACCTTCTGACCATCAGTTTCATCTTTCCTTTGATTGGCCATGTCACG 720
 || || || |||| ||||| ||||| | || |||| ||||| ||||| |||||
 Db 658 AACATAGATGTCACTCTTCTGGTCATCAGCTGTATTTTCCCTTGGTCGGCCATGTCATG 717

Qy 721 GGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCTTA 780
 || || ||||| ||||| | ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 718 GGCTTCTGCTGGCGTTCCTCACCCACCAGTCTTGGCAAAGGTGCAGGACGATTTCCATA 777

Qy 781 GAAACTGGAGCTCAGAATATTCAGATGTGCATCACCATGCTCCAGTTATCTTTCACTGCT 840
 || || |||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 778 GAGACCGGAGCACAGAACATCCAGCTGTGCATTGCCATGATGCAGCTGTCCTTCTCTGCT 837

Qy 841 GAGCACTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGAT 900
 |||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 838 GAGTACCTGGTCCAGCTGTTAAACTTCGCCCTGGCCTACGGACTCTTCCAAGTGTGCAC 897

Qy 901 GGATTTCTTATTGTTGCAGCATATCAGACGTACAAGAGGAGATTGAAGAACAACATGGA 960
 || | || |||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 898 GGGCTGCTCATTGTGCGCAGCATATCAGGCATACAAGAGGAGGCAGAAGAGTCAATACAGG 957

Qy 961 AAAAAGAACTCAGGTTGCACAGAAGTCTGCCATACGAGGAAATCGACTTCTTCCAGAGAG 1020
 | | || || || |||| |||| |||| |||| | |||| | ||||| |||||
 Db 958 AGACAGCACCCGGAGTGCCAAGACATCAGCTCTGAGAAGCA-----GCCAGAGAG 1008

Qy 1021 ACCAATGCCTTCTTGGAGGTGAATGAAGAAGGTGCCATCACTCCTGGGCCACCAGGGCCA 1080
 |||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1009 ACCAGTGCCTTCTTG-----GATAAAGGGGCTGAGGCTGCTGTAAGTCTGGGGCTA 1059

Qy 1081 ATGGATTGCCACAGGGCTCTCGAGCCAGTTGGCCACATCACTTCATGTGAATAG 1134
 | | ||||| | || | |||| | ||||| ||||| ||||| |||||
 Db 1060 GAGCAGCACACAGGACCGCTGAAGTACAGTCACGTTCTTCATGTGAATAG 1113

RESULT 5
 AX574600
 LOCUS AX574600 987 bp DNA linear PAT 07-JAN-2003
 DEFINITION Sequence 11 from Patent WO0233087.
 ACCESSION AX574600
 VERSION AX574600.1 GI:27551854
 KEYWORDS .
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS Edinger,S., Gerlach,V., Macdougall,J.R., Malyankar,U.M.,
 Smithson,G., Millet,I., Peyman,J.A., Stone,D.J., Gunther,E.,
 Ellerman,K., Shimkets,R.A., Padigar,M., Guo,X., Patturajan,M.,
 Taupier,R.J., Burgess,C.E., Zerhusen,B.D., Kekuda,R., Spytek,K.A.,
 Gangolli,E.A., Fernandes,E.R. and Gorman,L.
 TITLE Proteins and nucleic acids encoding same
 JOURNAL Patent: WO 0233087-A 11 25-APR-2002;
 Curagen Corporation (US)
 FEATURES Location/Qualifiers
 source 1. .987
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN

Query Match 57.8%; Score 655.8; DB 6; Length 987;
 Best Local Similarity 86.3%; Pred. No. 1.8e-189;
 Matches 803; Conservative 0; Mismatches 77; Indels 51; Gaps 5;

Qy	1	ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG	60
Db	1	ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG	60
Qy	61	CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTG	120
Db	61	CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTATC	120
Qy	121	ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCG	180
Db	121	ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCG	180
Qy	181	CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT	240
Db	181	CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT	240
Qy	241	TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT	300
Db	241	TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT	300
Qy	301	CTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGAT	360
Db	301	CTCATCATGGGCTGCTG-CCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGAT	359
Qy	361	GGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATG	420


```

Db      360 GGAGATATGGATCTCA-----GGTGCCCTGGGAATG 390
Qy      421 ATGCCACTCTGCATTTATCTCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATT 480
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      391 ATGCCACTCTGCATTTATCTCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATT 450
Qy      481 CCTTATCAGAACA-----TAGGAATTACCCTTGTGTGCCTGACCATTCTGTG 528
        ||||||||||||| |||||||||||||||||||||||||||||||||||
Db      451 CCTTATCAGAACATAGGTCTGTCTTTAGGAATTACCCTTGTGTGCCTGACCATTCTGTG 510
Qy      529 GCCTTTGGTGTCTATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATT 588
        |||||||||||||||||||||||||||||||||||||||||||||||
Db      511 GCCTTTGGTGTCTATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAA---- 566
Qy      589 GGGGCCGTTGTTGGTGGGGTCCTCCTTCTGGTGGTCGCAGTTGCTGGTGTGGTCTGGCG 648
        |||||||||||||||||||||||||||||||||||||||||||||||
Db      567 --GGCCGTTGTTGGTGGGGTCCTCCTTCTGGTGGTCGCAGTTGCTGGTGTGGTCTGGCG 624
Qy      649 AAAGGATCTTGGAATTACAGACATCACCTTCTGACCATCAGTTTCATCTTTCCTTTGATT 708
        |||||||||||||||||||||||||||||||||||||||||||||||
Db      625 AAAGGATCTTGGAATTACAGACATCACCTTCTGACCATCAGTTTCATCTTTCCTTTGATT 684
Qy      709 GGCCATGTCACGGGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGG 768
        ||||||||||||||||||||||||||||||||||||||||||||| |
Db      685 GGCCATGTCACGGGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGACCTTG 744
Qy      769 ACAATTTTCCTTAGAAACTGGAGCTCAGAATATTAGATGTGCATCACCATGCTCCAGTTA 828
        | | | | | | | | | | | | | | | | | | | | | | | |
Db      745 CCTATCTTTTGTAG---GTTTAGCTTTCAAGACACCCTGTGATACCCTACTCGCAATGACT 801
Qy      829 TCTTTCAGTCTGAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTC 888
        || | | | | | | | | | | | | | | | | | | | | |
Db      802 TCGTGTCTGAATGTTCCAGGCTCATCTATGCCTTCATTCCTCTGCTATATGGACTCTTC 861
Qy      889 CAGCTGATAGATGGATTTCTTATTGTTGCAG 919
        |||||||||||||||||||||||||||||
Db      862 CAGCTGATAGATGGATTTCTTATTGTTGAAG 892

```

RESULT 6

AC079237/c

LOCUS AC079237 23618 bp DNA linear PRI 21-FEB-2002

DEFINITION Homo sapiens BAC clone RP11-711J3 from 4, complete sequence.

ACCESSION AC079237

VERSION AC079237.7 GI:18482358

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 23618)

AUTHORS Sulston, J.E. and Waterston, R.

TITLE Toward a complete human genome sequence

JOURNAL Genome Res. 8 (11), 1097-1108 (1998)

MEDLINE 99063792

PUBMED 9847074

REFERENCE 2 (bases 1 to 23618)

AUTHORS Radionenko,M. and Meyer,R.
 TITLE The sequence of Homo sapiens BAC clone RP11-711J3
 JOURNAL Unpublished (2001)
 REFERENCE 3 (bases 1 to 23618)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (24-AUG-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 REFERENCE 4 (bases 1 to 23618)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (03-FEB-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 REFERENCE 5 (bases 1 to 23618)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (21-FEB-2002) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT On Feb 3, 2002 this sequence version replaced gi:18151062.

----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: sapiens@watson.wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0711J03

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by sequence
 from more than one subclone; and the assembly was confirmed by
 restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.
 McPherson, Department of Genetics, Washington University, St. Louis
 MO. For additional information about the map position of this
 sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male
 donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
 Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
 approach for construction of bacterial artificial chromosome
 libraries. Genomics 51:1-8. The clone may be obtained either from
 Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong
 and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-64A1, 2000 bp overlap; the clone sequenced to the right is RP11-397E7, 2000 bp overlap.

Actual start of this clone is at base position 1995 of RP11-64A1; actual end is at base position 108789 of RP11-397E7.

FEATURES	Location/Qualifiers
source	1. .23618 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="4" /map="4" /clone="RP11-711J3" /clone_lib="RPCI-11"
repeat_region	595. .720 /rpt_family="Alu"
repeat_region	789. .1014 /rpt_family="Alu"
repeat_region	1039. .1072 /rpt_family="(TAAA)n"
repeat_region	1734. .1936 /rpt_family="MIR"
repeat_region	2029. .2175 /rpt_family="L1"
misc_feature	2339. .2706 /note="similar to Sus scrofa EST BE031975 (NID:g8326984)"
misc_feature	2484. .2656 /note="match to EST BE181226 (NID:g8660402)"
misc_feature	2499. .2656 /note="similar to Mus musculus EST BB613812 (NID:g16454310)"
repeat_region	2740. .2892 /rpt_family="MIR"
repeat_region	3617. .3920 /rpt_family="Alu"
repeat_region	4602. .4890 /rpt_family="Alu"
repeat_region	4892. .4914 /rpt_family="AT_rich"
repeat_region	5459. .5624 /rpt_family="Alu"
misc_feature	6280. .6368 /note="match to EST BE181226 (NID:g8660402)"
misc_feature	6280. .6368 /note="similar to Mus musculus EST BB613812 (NID:g16454310)"
repeat_region	6811. .7098 /rpt_family="Alu"
repeat_region	7515. .7694 /rpt_family="MIR"
misc_feature	7794. .7912 /note="similar to Mus musculus EST BB613812 (NID:g16454310)"
misc_feature	7794. .7885 /note="match to EST BE181226 (NID:g8660402)"

repeat_region	8198. .8513	/rpt_family="L1"
repeat_region	8514. .8910	/rpt_family="MaLR"
repeat_region	8911. .9173	/rpt_family="L1"
repeat_region	9184. .9592	/rpt_family="MaLR"
repeat_region	9873. .10183	/rpt_family="Alu"
repeat_region	10157. .10266	/rpt_family="GA-rich"
repeat_region	10822. .11165	/rpt_family="MaLR"
repeat_region	11242. .11261	/rpt_family="(TTTTG)n"
repeat_region	11243. .11530	/rpt_family="Alu"
repeat_region	11690. .11833	/rpt_family="GA-rich"
repeat_region	11870. .11929	/rpt_family="L1"
repeat_region	11916. .11938	/rpt_family="AT-rich"
repeat_region	12632. .12696	/rpt_family="L1"
repeat_region	12842. .12935	/rpt_family="L2"
repeat_region	13129. .13424	/rpt_family="Alu"
repeat_region	13401. .13424	/rpt_family="(A)n"
repeat_region	13607. .13641	/rpt_family="L1"
repeat_region	13642. .13952	/rpt_family="Alu"
repeat_region	13953. .14359	/rpt_family="L1"
repeat_region	14392. .14416	/rpt_family="(T)n"
repeat_region	14562. .14588	/rpt_family="AT-rich"
repeat_region	15001. .15392	/rpt_family="L1"
repeat_region	16436. .16604	/rpt_family="MIR"
repeat_region	16895. .17214	/rpt_family="Alu"
repeat_region	17186. .17231	/rpt_family="(GAAAA)n"
repeat_region	17275. .17373	/rpt_family="(TTTC)n"
repeat_region	17344. .17657	/rpt_family="Alu"
repeat_region	17713. .17824	/rpt_family="L2"
repeat_region	18000. .18165	

```

repeat_region      /rpt_family="L1"
                    18255. .18384
repeat_region      /rpt_family="MIR"
                    18385. .18693
repeat_region      /rpt_family="Alu"
                    18666. .18695
repeat_region      /rpt_family="AT_rich"
                    18694. .18807
repeat_region      /rpt_family="MIR"
                    18881. .19188
repeat_region      /rpt_family="MaLR"
                    19189. .19562
repeat_region      /rpt_family="MaLR"

```

```

Query Match          33.2%;  Score 377;  DB 9;  Length 23618;
Best Local Similarity 100.0%;  Pred. No. 1.2e-103;
Matches 377;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

Qy      1 ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG 60
        |||
Db      23603 ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG 23544

Qy      61 CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTG 120
        |||
Db      23543 CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTG 23484

Qy      121 ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTTCG 180
        |||
Db      23483 ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTTCG 23424

Qy      181 CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT 240
        |||
Db      23423 CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT 23364

Qy      241 TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT 300
        |||
Db      23363 TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT 23304

Qy      301 CTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGAT 360
        |||
Db      23303 CTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGAT 23244

Qy      361 GGAGATATGGATCTCAG 377
        |||
Db      23243 GGAGATATGGATCTCAG 23227

```

RESULT 7

AC093827/c

```

LOCUS      AC093827          192263 bp    DNA    linear    PRI 01-MAR-2002
DEFINITION Homo sapiens BAC clone RP11-397E7 from 4, complete sequence.
ACCESSION  AC093827 AC016973
VERSION    AC093827.3  GI:16328304
KEYWORDS   HTG.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 192263)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074

REFERENCE 2 (bases 1 to 192263)
AUTHORS Goyea, E., Meyer, R. and Dixon, R.
TITLE The sequence of Homo sapiens BAC clone RP11-397E7
JOURNAL Unpublished (2001)

REFERENCE 3 (bases 1 to 192263)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE 4 (bases 1 to 192263)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE 5 (bases 1 to 192263)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

COMMENT On Oct 23, 2001 this sequence version replaced gi:15809171.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH0397E07
Drafting Center: WIBR
-----.

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-711J3; the clone sequenced to the right is RP11-168E22. Actual start of this clone is at base position 1 of RP11-397E7; actual end is at base position 192263 of RP11-397E7.

Data from AC079237 and AC093779 was used to finish this clone, AC093827. Polymorphisms have been identified between AC079237 and AC093827.

The sequence of AC016973 has been incorporated into AC093827.

FEATURES	Location/Qualifiers
source	1. .192263 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="4" /map="4" /clone="RP11-397E7" /clone_lib="RPCI-11"
repeat_region	1. .194 /rpt_family="MIR"
misc_feature	1553. .1820 /note="similar to EST BE151388 (NID:g8614109)"
repeat_region	3422. .3792 /rpt_family="MaLR"
repeat_region	3917. .3942 /rpt_family="(T)n"
repeat_region	4460. .4772 /rpt_family="Alu"
repeat_region	4750. .4772 /rpt_family="(A)n"
repeat_region	6150. .6469 /rpt_family="MaLR"
repeat_region	6470. .6781 /rpt_family="Alu"
repeat_region	6754. .6795 /rpt_family="(GAAAA)n"
repeat_region	6823. .7127 /rpt_family="Alu"
repeat_region	7245. .7324 /rpt_family="L1"
repeat_region	7386. .7651 /rpt_family="Alu"
repeat_region	7721. .7831 /rpt_family="L1"

```

repeat_region 7853. .8154
                /rpt_family="Alu"
repeat_region 7967. .7987
                /rpt_family="AT_rich"
repeat_region 8134. .8165
                /rpt_family="AT_rich"
repeat_region 8269. .8410
                /rpt_family="L1"
repeat_region 9108. .9235
                /rpt_family="MIR"
repeat_region 9966. .10007
                /rpt_family="MIR"
repeat_region 10729. .11025
                /rpt_family="Alu"
repeat_region 11003. .11025
                /rpt_family="(A)n"
repeat_region 11207. .11818
                /rpt_family="L1"
repeat_region 11808. .11850
                /rpt_family="AT_rich"
repeat_region 11851. .12151
                /rpt_family="Alu"
misc_feature 12109. .12705
                /note="similar to EST BG619594 (NID:g13670965)"
repeat_region 12132. .12151
                /rpt_family="(A)n"
repeat_region 12788. .12884
                /rpt_family="Alu"
repeat_region 13357. .13649
                /rpt_family="Alu"
repeat_region 13735. .14280
                /rpt_family="L1"
repeat_region 14376. .14499
                /rpt_family="Alu"
repeat_region 14500. .15875
                /rpt_family="L1"
repeat_region 15857. .15880
                /rpt_family="AT_rich"
repeat_region 15876. .16047
                /rpt_family="Alu"
repeat_region 16086. .16231
                /rpt_family="Alu"
repeat_region 16254. .16690
                /rpt_family="L2"
repeat_region 21991. .22029
                /rpt_family="AT_rich"
repeat_region 22069. .22413
                /rpt_family="L2"
repeat_region 22460. .22610
                /rpt_family="Alu"
misc_feature 23018. .23063
                /note="similar to EST BI522604 (NID:g15347396)"
misc_feature 23034. .23652
                /note="similar to EST BG707405 (NID:g13983721)"
misc_feature 23040. .23737
                /note="similar to EST BG612893 (NID:g13664264)"
misc_feature 23042. .23907

```


misc_feature /note="similar to EST BI549585 (NID:g15436897)"
 23042. .23852
 misc_feature /note="similar to EST BG779925 (NID:g14050242)"
 23042. .23791
 misc_feature /note="similar to EST BG401954 (NID:g13295402)"
 23042. .23696
 misc_feature /note="similar to EST BG720731 (NID:g13999918)"
 23042. .23491
 misc_feature /note="similar to EST BG530513 (NID:g13522050)"
 23043. .23763
 misc_feature /note="similar to EST BF700412 (NID:g11985820)"
 23043. .23273
 misc_feature /note="similar to EST BG771989 (NID:g14082642)"
 23044. .23864
 misc_feature /note="similar to EST BG528480 (NID:g13520017)"
 23053. .23874
 misc_feature /note="similar to EST BI838666 (NID:g15950216)"
 23055. .23873
 misc_feature /note="similar to EST BG822780 (NID:g14170367)"
 23056. .23907
 misc_feature /note="similar to EST BG680207 (NID:g13911604)"

Query Match 33.2%; Score 377; DB 9; Length 192263;
 Best Local Similarity 100.0%; Pred. No. 1.6e-103;
 Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1985 ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG 1926
 Qy 61 CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTG 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1925 CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTG 1866
 Qy 121 ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCG 180
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1865 ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCG 1806
 Qy 181 CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT 240
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1805 CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT 1746
 Qy 241 TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT 300
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1745 TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT 1686
 Qy 301 CTCATCATGGGCTGCTGCCCCGGGGGACCATCTCTAACATTTTCACCTTCTGGGTTGAT 360
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1685 CTCATCATGGGCTGCTGCCCCGGGGGACCATCTCTAACATTTTCACCTTCTGGGTTGAT 1626
 Qy 361 GGAGATATGGATCTCAG 377
 ||||||||||||||||
 Db 1625 GGAGATATGGATCTCAG 1609

LOCUS AC099847 65268 bp DNA linear HTG 22-NOV-2001
 DEFINITION Homo sapiens chromosome 18 clone RP11-819K4 map 18, LOW-PASS
 SEQUENCE SAMPLING.
 ACCESSION AC099847
 VERSION AC099847.1 GI:17047210
 KEYWORDS HTG; HTGS_PHASE0.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 65268)
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 TITLE Homo sapiens chromosome 18, clone RP11-819K4
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 65268)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
 Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
 Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
 Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
 MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
 Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 TITLE Direct Submission
 JOURNAL Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L13211
 Center clone name: 819_K_4

 * NOTE: This record contains 81 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for

* identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

*	1	727: contig of 727 bp in length
*	728	827: gap of 100 bp
*	828	1547: contig of 720 bp in length
*	1548	1647: gap of 100 bp
*	1648	2369: contig of 722 bp in length
*	2370	2469: gap of 100 bp
*	2470	3148: contig of 679 bp in length
*	3149	3248: gap of 100 bp
*	3249	3922: contig of 674 bp in length
*	3923	4022: gap of 100 bp
*	4023	4708: contig of 686 bp in length
*	4709	4808: gap of 100 bp
*	4809	5512: contig of 704 bp in length
*	5513	5612: gap of 100 bp
*	5613	6320: contig of 708 bp in length
*	6321	6420: gap of 100 bp
*	6421	7119: contig of 699 bp in length
*	7120	7219: gap of 100 bp
*	7220	7943: contig of 724 bp in length
*	7944	8043: gap of 100 bp
*	8044	8780: contig of 737 bp in length
*	8781	8880: gap of 100 bp
*	8881	9602: contig of 722 bp in length
*	9603	9702: gap of 100 bp
*	9703	10429: contig of 727 bp in length
*	10430	10529: gap of 100 bp
*	10530	11235: contig of 706 bp in length
*	11236	11335: gap of 100 bp
*	11336	12030: contig of 695 bp in length
*	12031	12130: gap of 100 bp
*	12131	12822: contig of 692 bp in length
*	12823	12922: gap of 100 bp
*	12923	13644: contig of 722 bp in length
*	13645	13744: gap of 100 bp
*	13745	14466: contig of 722 bp in length
*	14467	14566: gap of 100 bp
*	14567	15297: contig of 731 bp in length
*	15298	15397: gap of 100 bp
*	15398	16109: contig of 712 bp in length
*	16110	16209: gap of 100 bp
*	16210	16935: contig of 726 bp in length
*	16936	17035: gap of 100 bp
*	17036	17749: contig of 714 bp in length
*	17750	17849: gap of 100 bp
*	17850	18542: contig of 693 bp in length
*	18543	18642: gap of 100 bp
*	18643	19341: contig of 699 bp in length
*	19342	19441: gap of 100 bp
*	19442	20144: contig of 703 bp in length
*	20145	20244: gap of 100 bp
*	20245	20963: contig of 719 bp in length

*	20964	21063: gap of 100 bp
*	21064	21779: contig of 716 bp in length
*	21780	21879: gap of 100 bp
*	21880	22596: contig of 717 bp in length
*	22597	22696: gap of 100 bp
*	22697	23361: contig of 665 bp in length
*	23362	23461: gap of 100 bp
*	23462	24186: contig of 725 bp in length
*	24187	24286: gap of 100 bp
*	24287	25016: contig of 730 bp in length
*	25017	25116: gap of 100 bp
*	25117	25835: contig of 719 bp in length
*	25836	25935: gap of 100 bp
*	25936	26607: contig of 672 bp in length
*	26608	26707: gap of 100 bp
*	26708	27297: contig of 590 bp in length
*	27298	27397: gap of 100 bp
*	27398	28099: contig of 702 bp in length
*	28100	28199: gap of 100 bp
*	28200	28896: contig of 697 bp in length
*	28897	28996: gap of 100 bp
*	28997	29707: contig of 711 bp in length
*	29708	29807: gap of 100 bp
*	29808	30519: contig of 712 bp in length
*	30520	30619: gap of 100 bp
*	30620	31348: contig of 729 bp in length
*	31349	31448: gap of 100 bp
*	31449	32177: contig of 729 bp in length
*	32178	32277: gap of 100 bp
*	32278	33007: contig of 730 bp in length
*	33008	33107: gap of 100 bp
*	33108	33808: contig of 701 bp in length
*	33809	33908: gap of 100 bp
*	33909	34618: contig of 710 bp in length
*	34619	34718: gap of 100 bp
*	34719	35387: contig of 669 bp in length
*	35388	35487: gap of 100 bp
*	35488	36158: contig of 671 bp in length
*	36159	36258: gap of 100 bp
*	36259	36957: contig of 699 bp in length
*	36958	37057: gap of 100 bp
*	37058	37764: contig of 707 bp in length
*	37765	37864: gap of 100 bp
*	37865	38561: contig of 697 bp in length
*	38562	38661: gap of 100 bp
*	38662	39375: contig of 714 bp in length
*	39376	39475: gap of 100 bp
*	39476	40191: contig of 716 bp in length
*	40192	40291: gap of 100 bp
*	40292	41017: contig of 726 bp in length
*	41018	41117: gap of 100 bp
*	41118	41764: contig of 647 bp in length
*	41765	41864: gap of 100 bp
*	41865	42593: contig of 729 bp in length
*	42594	42693: gap of 100 bp
*	42694	43356: contig of 663 bp in length
*	43357	43456: gap of 100 bp

* 43457 44158: contig of 702 bp in length
 * 44159 44258: gap of 100 bp
 * 44259 44951: contig of 693 bp in length
 * 44952 45051: gap of 100 bp
 * 45052 45766: contig of 715 bp in length
 * 45767 45866: gap of 100 bp
 * 45867 46581: contig of 715 bp in length
 * 46582 46681: gap of 100 bp
 * 46682 47400: contig of 719 bp in length
 * 47401 47500: gap of 100 bp
 * 47501 48234: contig of 734 bp in length
 * 48235 48334: gap of 100 bp
 * 48335 49057: contig of 723 bp in length
 * 49058 49157: gap of 100 bp
 * 49158 49877: contig of 720 bp in length
 * 49878 49977: gap of 100 bp
 * 49978 50699: contig of 722 bp in length
 * 50700 50799: gap of 100 bp
 * 50800 51494: contig of 695 bp in length
 * 51495 51594: gap of 100 bp
 * 51595 52298: contig of 704 bp in length
 * 52299 52398: gap of 100 bp
 * 52399 53096: contig of 698 bp in length
 * 53097 53196: gap of 100 bp
 * 53197 53903: contig of 707 bp in length
 * 53904 54003: gap of 100 bp
 * 54004 54708: contig of 705 bp in length
 * 54709 54808: gap of 100 bp
 * 54809 55524: contig of 716 bp in length

Query Match 33.1%; Score 375.4; DB 2; Length 65268;
 Best Local Similarity 99.7%; Pred. No. 4.3e-103;
 Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG 60
 ||||||| ||||||||||||||||||||||||||||||||||||||||||||
 Db 50030 ATGAGAGCCCATTTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG 50089

 Qy 61 CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTG 120
 ||||||| ||||||||||||||||||||||||||||||||||||||||||||
 Db 50090 CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTG 50149

 Qy 121 ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCTG 180
 ||||||| ||||||||||||||||||||||||||||||||||||||||||||
 Db 50150 ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCTG 50209

 Qy 181 CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT 240
 ||||||| ||||||||||||||||||||||||||||||||||||||||||||
 Db 50210 CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT 50269

 Qy 241 TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT 300
 ||||||| ||||||||||||||||||||||||||||||||||||||||||||
 Db 50270 TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT 50329

 Qy 301 CTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGAT 360
 ||||||| ||||||||||||||||||||||||||||||||||||||||||||
 Db 50330 CTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGAT 50389

```

Qy      361 GGAGATATGGATCTCAG 377
          |||||
Db      50390 GGAGATATGGATCTCAG 50406

```

RESULT 9

AR033870

LOCUS	AR033870	2263 bp	DNA	linear	PAT 29-SEP-1999
-------	----------	---------	-----	--------	-----------------

DEFINITION Sequence 1 from patent US 5869265.

ACCESSION AR033870

VERSION AR033870.1 GI:5949475

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

Unclassified.

REFERENCE 1 (bases 1 to 2263)

AUTHORS Dawson, P.A.

TITLE	Ileal bile acid transporter compositions and methods
-------	--

JOURNAL Patent: US 5869265-A 1 09-FEB-1999;

FEATURES	Location/Qualifiers
----------	---------------------

```
source      1.  .2263
```

```
/organism="unknown"
```

```
/mol type="unassigned DNA"
```

ORIGIN

Query Match 28.3%; Score 320.4; DB 6; Length 2263;

Best Local Similarity 60.8%; Pred. No. 2.1e-86;

Matches 522; Conservative 0; Mismatches 336; Indels 0; Gaps 0;

Qy 80 ATGGAACCTGGAGCTCGTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGT 139

| | | | |

Db 188 ACGCCATCCTCAGCGTGGTGATGAGCACCGTGCTCACAATCCTCCTAGCCTTGGTGATGT 247

Qy 140 TCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCGCACATCAGGAGACCCTGGG 199

| | | | | | | | | | | | | | | | | | | | | |

Db 248 TTTCCATGGGGTGCAATGTGGAAGTCCACAAGTTTCTGGGACACCTAAGGCGGCCATGGG 307

Qv 200 GCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGG 259

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 308 GCATCGTCGTGGGCTTCCTCTGTCAGTTTGGAATCATGCCTCTCACAGGTTTCGTCCTGT 367

Qy 260 CCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCC 319

|||||

Db 368 CCGTGGCCTTTGGCATCCTCCCAGTGCAAGCTGTGGTGGTGCTGATCCAGGGTTGCTGCC 427

Qv 320 CGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGATGGAGATATGGATCTCAGCA 379

Db 428 CTGGAGGAACTGCCTCCAATATCCTAGCCTATTGGGTAGATGGCGACATGGACCTCAGCG 487

Qv 380 TCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATC 439

Db 488 TTAGCATGACCACCTGCTCCACGCTGCTTGCCCTTGAATGATGCCCCCTTGCCTCTTCA 547

Qv 440 TCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATTCTTATCAGAACATAGGAA 499

||||| ||||| | ||||| | || | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 548 TCTATACCAAGATGTGGGTTGACTCAGGGACGATTGTGATTCCTTATGACAGCATTGGCA 607

Qy 500 TTACCCCTTGTGTGCCTGACCATTCTGTGGCCTTTGGTGTCTATGTGAATTACAGATGGC 559
 | | | | | | | | | | | | | | | | | | | | | |
 Db 608 CTTCTCTGGTTGCTCTTGTATTCTGTTCCATTGGAATGTATGTGAATCACAAATGGC 667

Qy 560 CAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTTGGTGGGGTCCTCCTTCTGG 619
 | | | | | | | | | | | | | | | | | | | | | |
 Db 668 CCCAAAAGCAAAGATCATACTTAAAATTGGATCCATCGCAGGTGCAATTCTCATTGTTC 727

Qy 620 TGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGATCTTGAATTCAGACATCACCCCTTC 679
 | | | | | | | | | | | | | | | | | | | | | |
 Db 728 TCATCGCTGTGGTTGGAGGAATACTGTACCAAAGTGCCTGGACCATTGAACCCAAGCTGT 787

Qy 680 TGACCATCAGTTTTCATCTTTCTTTGATTGGCCATGTCACGGGTTTTCTGCTGGCACTTT 739
 | | | | | | | | | | | | | | | | | | | | | |
 Db 788 GGATTATAGGAACCATATATCTATAGCTGGCTACGGCCTGGGGTTTTCTGGCTAGAA 847

Qy 740 TTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCTTAGAAACTGGAGCTCAGAATA 799
 | | | | | | | | | | | | | | | | | | | | | |
 Db 848 TTGCTGGTCAACCCTGGTACAGGTGCCGAACAGTTGCCTTGGAACCGGGTTGCAGAACA 907

Qy 800 TTCAGATGTGCATCACCATGCTCCAGTTATCTTTCAGTGTGAGCACTTGGTCCAGATGT 859
 | | | | | | | | | | | | | | | | | | | | | |
 Db 908 CTCAGCTGTGTTCCACCATTGTGCAGCTTTCCTTCAGCCCTGAGGACCTCAACCTTGTGT 967

Qy 860 TGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGATGGATTTCTTATTGTTGCAG 919
 | | | | | | | | | | | | | | | | | | | | | |
 Db 968 TCACCTTCCCCCTCATCTACAGCATCTTCCAGATCGCCTTTGCAGCAATACTATTAGGAG 1027

Qy 920 CATATCAGACGTACAAGA 937
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1028 CTTATGTCGCATACAAGA 1045

RESULT 10

I32744

LOCUS I32744 2263 bp DNA linear PAT 06-FEB-1997

DEFINITION Sequence 1 from patent US 5589358.

ACCESSION I32744

VERSION I32744.1 GI:1823535

KEYWORDS .

SOURCE Unknown.

ORGANISM Unknown.

Unclassified.

REFERENCE 1 (bases 1 to 2263)

AUTHORS Dawson,P.A.

TITLE Ileal bile acid transporter compositions and methods

JOURNAL Patent: US 5589358-A 1 31-DEC-1996;

FEATURES Location/Qualifiers

source 1..2263

/organism="unknown"

/mol_type="unassigned DNA"

ORIGIN

Query Match 28.3%; Score 320.4; DB 6; Length 2263;

Best Local Similarity 60.8%; Pred. No. 2.1e-86;

Matches 522; Conservative 0; Mismatches 336; Indels 0; Gaps 0;

Qy 80 ATGGAACCTGGAGCTCGTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGT 139
 | | | | | | | | | | | | | | | | | | | | | |
 Db 188 ACGCCATCCTCAGCGTGGTGTGAGCACCGTGTCTACAATCCTCCTAGCCTTGGTGATGT 247

Qy 140 TCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCGCACATCAGGAGACCCTGGG 199
 | | | | | | | | | | | | | | | | | | | | | |
 Db 248 TTTCCATGGGGTGCAATGTGGAACCTCACAAAGTTTCTGGGACACCTAAGGCGGCCATGGG 307

Qy 200 GCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGG 259
 | | | | | | | | | | | | | | | | | | | | | |
 Db 308 GCATCGTCGTGGGCTTCTCTGTCTAGTTTGAATCATGCCTCTCACAGTTTTCGTCTGT 367

Qy 260 CCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCC 319
 | | | | | | | | | | | | | | | | | | | | | |
 Db 368 CCGTGGCCTTTGGCATCCTCCAGTGCAAGCTGTGGTGGTGTGATCCAGGGTTGCTGCC 427

Qy 320 CGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGATGGAGATATGGATCTCAGCA 379
 | | | | | | | | | | | | | | | | | | | | | |
 Db 428 CTGGAGGAACTGCCTCCAATATCCTAGCCTATTGGGTAGATGGCGACATGGACCTCAGCG 487

Qy 380 TCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATC 439
 | | | | | | | | | | | | | | | | | | | | | |
 Db 488 TTAGCATGACCACCTGCTCCACGCTGCTTGCCCTTGGAATGATGCCCTTGCCTCTTCA 547

Qy 440 TCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATTCTTATCAGAACATAGGAA 499
 | | | | | | | | | | | | | | | | | | | | | |
 Db 548 TCTATACCAAGATGTGGGTTGACTCAGGGACGATTGTGATTCTTATGACAGCATTGGCA 607

Qy 500 TTACCCTTGTGTGCCTGACCATTCTGTGGCCTTTGGTGTCTATGTGAATTACAGATGGC 559
 | | | | | | | | | | | | | | | | | | | | | |
 Db 608 CTTCTCTGGTTGCTCTTGTTATTCTGTTCATTGGAATGTATGTGAATCACAAATGGC 667

Qy 560 CAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTTGGTGGGGTCTCCTTCTGG 619
 | | | | | | | | | | | | | | | | | | | | | |
 Db 668 CCCAAAAGCAAAGATCATACTTAAATTTGGATCCATCGCAGGTGCAATTCTCATTGTTC 727

Qy 620 TGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGATCTTGAATTCAGACATCACCCCTC 679
 | | | | | | | | | | | | | | | | | | | | | |
 Db 728 TCATCGCTGTGGTTGGAGGAATACTGTACCAAAGTGCCTGGACCATGAACCCAAGCTGT 787

Qy 680 TGACCATCAGTTTCATCTTTTCTTTGATTGGCCATGTCACGGGTTTTCTGCTGGCACTTT 739
 | | | | | | | | | | | | | | | | | | | | | |
 Db 788 GGATTATAGGAACCATATATCCTATAGCTGGCTACGGCCTGGGGTTTTTCTGGCTAGAA 847

Qy 740 TTACCCACCACTCTTGGCAAAGGTGCAGGACAATTTCTTAGAACTGGAGCTCAGAATA 799
 | | | | | | | | | | | | | | | | | | | | | |
 Db 848 TTGCTGGTCAACCCTGGTACAGGTGCCGAACAGTTGCCTTGGAACCGGGTTGCAGAAC 907

Qy 800 TTCAGATGTGCATCACCATGCTCCAGTTATCTTTCACTGCTGAGCACTTGGTCCAGATGT 859
 | | | | | | | | | | | | | | | | | | | | | |
 Db 908 CTCAGCTGTGTTCCACCATTGTGCAGCTTTCTTCAGCCCTGAGGACCTCAACCTTGTGT 967

Qy 860 TGAGTTTCCCACTGGCCTATGGACTCTTCAGCTGATAGATGGATTCTTATTGTTGCAG 919
 | | | | | | | | | | | | | | | | | | | | | |
 Db 968 TCACCTTCCCCCTCATCTACAGCATCTTCAGATCGCCTTTGCAGCAATACTATTAGGAG 1027

Qy 920 CATATCAGACGTACAAGA 937
| | | | | | | |
Db 1028 CTTATGTCGCATACAAGA 1045

RESULT 11

CGU02028

LOCUS CGU02028 2263 bp mRNA linear ROD 06-JUN-1994

DEFINITION Cricetulus griseus Na⁺ dependent ileal bile acid transporter mRNA,
complete cds.

ACCESSION U02028

VERSION U02028.1 GI:455032

KEYWORDS .

SOURCE Cricetulus griseus (Chinese hamster)

ORGANISM Cricetulus griseus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.

REFERENCE 1 (bases 1 to 2263)

AUTHORS Wong, M.H., Oelkers, P., Craddock, A.L. and Dawson, P.A.

TITLE Expression cloning and characterization of the hamster ileal
sodium-dependent bile acid transporter

JOURNAL J. Biol. Chem. 269 (2), 1340-1347 (1994)

MEDLINE 94117449

PUBMED 8288599

REFERENCE 2 (bases 1 to 2263)

AUTHORS Dawson, P.A.

TITLE Direct Submission

JOURNAL Submitted (22-SEP-1993) Paul A. Dawson, Dept Medicine/Section
Gastroenterology, Bowman Gray School of Medicine, Wake Forest
University, Medical Center Boulevard, Winston-Salem, NC, 27517, USA

FEATURES Location/Qualifiers

source

1. .2263

/organism="Cricetulus griseus"

/mol_type="mRNA"

/db_xref="taxon:10029"

/clone="clone pIBAT(44-1)"

/tissue_type="ileum"

/clone_lib="hamster ileal cDNA expression library"

/note="author cites additional common name: golden Syrian
hamster"

CDS

109. .1155

/codon_start=1

/product="Na⁺ dependent ileal bile acid transporter"

/protein_id="AAA18640.1"

/db_xref="GI:455033"

/translation="MDNSSICNP NATICEGDSCIAPESNFNAILSVMSTVL TILLAL
VMFSMGCNVELHKFLGHLRRPWGIVVGFLCQFGIMPLTG FVLSVAFGILPVQAVVLI
QGCCPGGTASNILAYWVDGMDLSVSM TTCSTLLALGMMPLCLFIYTKMWVDSGTIVI
PYDSIGTSLVALVIPV SIGMYVNHKWPQKAKIILKIGSIAGAILIVLIAVVG GILYQS
AWTIEPKLWIIGTIYPIAGYGLGFFLARIAGQPWYRCRTVALETGLQNTQLCSTIVQL
SFSPEDLNLVFTFPLIYSIFQIAFAA ILLGAYVAYKKCHGKNNTELQEKTDNEMEPRS
SFQETNKGFPDEK"

mat_peptide

109. .1152

/product="Na⁺ dependent ileal bile acid transporter"

ORIGIN

Query Match 28.3%; Score 320.4; DB 10; Length 2263;
Best Local Similarity 60.8%; Pred. No. 2.1e-86;
Matches 522; Conservative 0; Mismatches 336; Indels 0; Gaps 0;

```
Qy      80 ATGGAACCTGGAGCTCGTTTTACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGT 139
      | | | |||      | | | | |      ||| ||| | | | | | ||| ||||
Db      188 ACGCCATCCTCAGCGTGGTGATGAGCACCGTGCTCACAATCCTCCTAGCCTTGGTGATGT 247

Qy      140 TCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCGCACATCAGGAGACCTGGG 199
      | | | |||| | |      |||| | | | | | | | | | | | | | | | | |
Db      248 TTTCCATGGGGTGCAATGTGGAATCCACAAGTTTCTGGGACACCTAAGCGGCCATGGG 307

Qy      200 GCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGG 259
      |||| | | |||| | |||| |||| |||| | | | | | | | | | | | |
Db      308 GCATCGTCGTGGGCTTCCTCTGTCTAGTTTGAATCATGCCTCTCACAGGTTTCGTCCTGT 367

Qy      260 CCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCC 319
      || | | |||| | |      |||| |||| |||| | | | | | | | | | | | |
Db      368 CCGTGGCCTTTGGCATCCTCCAGTGCAAGCTGTGGTGGTGCTGATCCAGGGTTGCTGCC 427

Qy      320 CGGGGGGCACCATCTCTAACATTTTACCTTCTGGGTTGATGGAGATATGGATCTCAGCA 379
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      428 CTGGAGGAAGTGCCTCCAATATCCTAGCCTATTGGGTAGATGGCGACATGGACCTCAGCG 487

Qy      380 TCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATC 439
      | | | |||| | |||| |||| | | | | | | | | | | | | | | | |
Db      488 TTAGCATGACCACCTGCTCCACGCTGCTTGCCCTTGAATGATGCCCTTTGCCTCTTCA 547

Qy      440 TCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATTCTTATCAGAACATAGGAA 499
      |||| ||| | | ||| | |      | | | | | | | | | | | | | | | |
Db      548 TCTATACCAAGATGTGGGTTGACTCAGGGACGATTGTGATTCCTTATGACAGCATTGGCA 607

Qy      500 TTACCCTTGTGTGCCTGACCATTCTGTGGCCTTTGGTGTCTATGTGAATTACAGATGGC 559
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      608 CTTCTCTGGTTGCTCTTGTTATTCTGTTCATTGGAATGTATGTGAATCACAAATGGC 667

Qy      560 CAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTTGGTGGGGTCCTCCTTCTGG 619
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      668 CCCAAAAGCAAAGATCATACTTAAATTTGGATCCATCGCAGGTGCAATTCTCATTGTTC 727

Qy      620 TGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGATCTTGAATTCAGACATCACCTTC 679
      | | |||| | | | | | | | | | | | | | | | | | | | | | | |
Db      728 TCATCGCTGTGGTTGGAGGAATACTGTACCAAAGTGCCTGGACCATTGAACCAAGCTGT 787

Qy      680 TGACCATCAGTTTCATCTTTCTTTGATTGGCCATGTCACGGGTTTTCTGCTGGCACTTT 739
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      788 GGATTATAGGAACCATATATCCTATAGCTGGCTACGGCCTGGGGTTTTCTGGCTAGAA 847

Qy      740 TTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCTTAGAACTGGAGCTCAGAATA 799
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      848 TTGCTGGTCAACCCTGGTACAGGTGCCGAACAGTTGCCTTGGAACCGGGTTGCAGAACA 907

Qy      800 TTCAGATGTGCATCACCATGCTCCAGTTATCTTTCACTGCTGAGCACTTGGTCCAGATGT 859
      |||| |||| | |||| | ||| | | |||| | |||| | | | | | | | |
Db      908 CTCAGCTGTGTTCCACCATGTGTCAGCTTTCTTCAGCCCTGAGGACCTCAACCTTGTGT 967

Qy      860 TGAGTTTCCCACTGGCCTATGGACTCTTCAGCTGATAGATGGATTTCCTTATTGTTGCAG 919
```

Db 968 TCACCTCCCCCTCATCTACAGCATCTTCCAGATCGCCTTTGCAGCAATACTATTAGGAG 1027
 Qy 920 CATATCAGACGTACAAGA 937
 Db 1028 CTTATGTCGCATACAAGA 1045

RESULT 12

BC053189

LOCUS BC053189 1916 bp mRNA linear VRT 07-OCT-2003

DEFINITION Danio rerio cDNA clone MGC:63998 IMAGE:6792624, complete cds.

ACCESSION BC053189

VERSION BC053189.1 GI:31418837

KEYWORDS MGC.

SOURCE Danio rerio (zebrafish)

ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.

REFERENCE 1 (bases 1 to 1916)

AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
Schnierch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

TITLE Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

MEDLINE 22388257

PUBMED 12477932

REFERENCE 2 (bases 1 to 1916)

AUTHORS Strausberg,R.

TITLE Direct Submission

JOURNAL Submitted (02-JUN-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Leonard I. Zon, M.D.

cDNA Library Preparation: Invitrogen Corp

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 117 Row: g Column: 20
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.

```
FEATURES             Location/Qualifiers
    source             1. .1916
                        /organism="Danio rerio"
                        /mol_type="mRNA"
                        /db_xref="taxon:7955"
                        /clone="MGC:63998 IMAGE:6792624"
                        /tissue_type="Kidney, zebrafish"
                        /clone_lib="NCI_CGAP_ZKid1"
                        /lab_host="DH10B"
                        /note="Vector: pCMV-SPORT6.1"
    CDS                168. .1253
                        /codon_start=1
                        /product="Unknown (protein for MGC:63998)"
                        /protein_id="AAH53189.1"
                        /db_xref="GI:31418838"
                        /translation="MCTLEPVCVPVNATICGTGTSCLVPRDPFNDILSVMSVAITVMLA
MVMFSMGCTVEARKLWGHVRRPWGIFIGFLCQFGIMPFTAFILSLLFENVLPVQAVVII
IMGCCPGGSSSNVFCYWLDGMDLSISMTACSSILALGMMPLCLLIYTTIWTAGDAIQ
IPYDNIGITLVSLLPVGLGMLVKHKWPKAARKILKVGSVVGIVLIIVIAVIGGVLYQ
SSWTIAPSLWIIIGTIYPPFIGFLGFLARFVGQPWHRCRTIALETGMQNAQLASTITQ
LSFSPAELVMFAFPLIYSIFQLVVAGIAVSIHYSIKRCRHQTLVEEDGEGTTEDCDK
HSYSLENGGFSCDENNGNNQNKDKGTKL"
    misc_feature       285. .731
                        /note="SBF; Region: Sodium Bile acid symporter family.
This family consists of Na+/bile acid co-transporters.
These transmembrane proteins function in the liver in the
uptake of bile acids from portal blood plasma a process
mediated by the co-transport of Na+. Also in the family is
ARC3 from S. cerevisiae, a putative transmembrane protein
involved in resistance to arsenic compounds"
                        /db_xref="CDD:pfam01758"
```

ORIGIN

Query Match 27.3%; Score 309.6; DB 5; Length 1916;
 Best Local Similarity 60.4%; Pred. No. 4.2e-83;
 Matches 510; Conservative 0; Mismatches 334; Indels 0; Gaps 0;

```
Qy      95 TCGT'TTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGTTCTCTTTGGGATGTT 154
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      265 TTGTGATGAGCGTTGCCATTACCGTCATGTTGGCCATGTTTATGTTTTC AATGGGCTGCA 324

Qy      155 CCGTGGAGATCCGGAAGCTGTGGTCGCACATCAGGAGACCCTGGGGCATTTGCTGTGGGAC 214
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      325 CTGTTGAGGCTAGAAAACGTGTGGGGGCACGTTTCGACAGCCCTGGGGCATTTTATAGGTT 384
```


DEFINITION House mouse; Musculus domesticus mRNA for ileal Na+-dependent bile acid transporter, partial cds.

ACCESSION D87059

VERSION D87059.1 GI:1504059

KEYWORDS ileal Na+-dependent bile acid transporter.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 974)

AUTHORS Saeki,T.

TITLE Mouse ileal Na-dependent bile acid transporter cDNA: Partial CDS

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 974)

AUTHORS Saeki,T.

TITLE Direct Submission

JOURNAL Submitted (09-AUG-1996) Tohru Saeki, Kyoto Prefectural University, Department of Biological Resource Chemistry; Nakaragi, Shimogamo, Sakyo-ku, Kyoto, Kyoto 606, Japan (E-mail:tsaeki@dns.kpu.ac.jp, Tel:81-75-703-5663, Fax:81-75-703-5661)

FEATURES Location/Qualifiers

source 1..974
/organism="Mus musculus"
/mol_type="mRNA"
/strain="ICR"
/db_xref="taxon:10090"
/dev_stage="8 weeks"

CDS <1..>974
/codon_start=1
/product="ileal Na+-dependent bile acid transporter"
/protein_id="BAA13237.1"
/db_xref="GI:1504060"
/translation="PNATVCEGDS CVVPESN FNAILNTVMSTVLTILLAMVMFSMGCN
VEVHKFLGHIKRPWGIFVGF L CQFGIMPLTGFI LSVASGILPVQAVVVLIMGCCPGGT
GSNILAYWIDGDM DLSVSM T TCSTLLALGMMPLCLFVYTKMWVDSGTIVIPYDSIGIS
LVALVIPVSFGMFVN HKWPQAKI I LKIGSITGVILIVLIAVIGGILYQSAWII EPKL
WIIGTIFPIAGYSLGFF LARLAGQPWYRCRTVALETGMQNTQLCSTIVQLSFS PEDLN
LVFTFPLIYTVFQLVFAAVILGIYV TYRKCYGKNDAEFLEKTDNEMDS"

ORIGIN

Query Match 27.1%; Score 307.8; DB 10; Length 974;
Best Local Similarity 59.4%; Pred. No. 1.4e-82;
Matches 522; Conservative 0; Mismatches 357; Indels 0; Gaps 0;

```

Qy      80 ATGGAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGT 139
      ||| || || | || | || | || | || | || | || | |||
Db      56 ATGCAATTCTCAATACAGTGATGAGCACTGTGCTCACCATCCTCTTAGCCATGGTGATGT 115

Qy     140 TCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCGCACATCAGGAGACCCTGGG 199
      | ||| |||| || ||||| ||| ||| | || || | ||||| ||||
Db     116 TTTCTATGGGTGCAATGTGGAAGTCCACAAGTTCCTAGGACATATAAAGAGACCATGGG 175

Qy     200 GCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGG 259
      | || |||| | |||| ||||| ||||| |||| | || |||
Db     176 GTATCTTCGTGGGCTTCCTCTGTCAAGTTTGAATCATGCCTCTCACAGGCTTTATCCTGT 235

Qy     260 CCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCC 319

```

```

      | |   || |   |   || || || || | |   || || || || || || || || || ||
Db      236 CTGTGGCCTCTGGCATCCTTCCTGTACAGGCTGTAGTGGTGCTAATTATGGGTTGCTGCC 295

Qy      320 CGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGATGGAGATATGGATCTCAGCA 379
      | || || || ||   ||| || || |   |||   ||| | |||| | | |||| | ||||
Db      296 CTGGAGGAACCTGGCTCCAATATCCTGGCCTATTGGATAGATGGCGACATGGACCTCAGTG 355

Qy      380 TCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATC 439
      | || |||| | | || |||| | |   |||| | |||| |||| || || || | |
Db      356 TTAGCATGACCACTTGCTCCACACTGCTTGCCCTTGGAATGATGCCTCTTTGCCTCTTCG 415

Qy      440 TCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATTCCCTTATCAGAACATAGGAA 499
      |||| || | | || | |   | |   | |   |||| || | | || | | | |
Db      416 TCTACACCAAGATGTGGGTTGACTCGGGAACGATTGTGATTCCCTATGATAGCATTGGTA 475

Qy      500 TTACCCTTGTGTGCCTGACCATTCCCTGTGGCCTTTGGTGTCTATGTGAATTACAGATGGC 559
      || | || ||   ||   |||| || | |||| | | || | || | || | || |
Db      476 TTTCTCTGGTTGCTCTTGTATTCCCTGTTCCCTTTGGAATGTTTGTAATCACAAATGGC 535

Qy      560 CAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTTGGTGGGGTCCTCCTTCTGG 619
      || || || | || || | || || || || || || |   || | | || | || |
Db      536 CACAAAAGCGAAGATTATACTTAAATTTGGATCCATCACAGGTGTAATTCTCATTGTGC 595

Qy      620 TGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGATCTTGAATTCAGACATCACCTTC 679
      | | || ||   || | |   || | |   || | |   || | |   || | |
Db      596 TCATAGCTGTGATTGGAGGAATACTGTACCAAAGTGCCTGGATCATTGAACCCAACTGT 655

Qy      680 TGACCATCAGTTTCATCTTTCCTTTGATTGGCCATGTCACGGGTTTTCTGCTGGCACTTT 739
      || || | |   || || || | || | |   || | |   || || | || || |
Db      656 GGATTATAGGAACAATATTCCTATAGCTGGCTACAGCCTGGGTTTTCTTCCTGGCTAGAC 715

Qy      740 TTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCTTAGAAACTGGAGCTCAGAATA 799
      | |   || | || | |||| | || | |   || | |||| || || || |
Db      716 TAGCTGGTCAACCCTGGTACAGGTGCCGAACAGTAGCCTTGGAAGCTGGAATGCAGAACA 775

Qy      800 TTCAGATGTGCATCACCATGCTCCAGTTATCTTTCAGTGTGAGCACTTGGTCCAGATGT 859
      |||| |||| | |||| | || | || || | | || | | | || | || |
Db      776 CTCAGCTGTGCTCCACCATGTACAGCTCTCCTTCTCCCCGAGGATCTCAACCTGGTGT 835

Qy      860 TGAGTTTCCCACTGGCCTATGGACTCTCCAGCTGATAGATGGATTTCTTATTGTTGCAG 919
      | | |||| || | || | | |||| || | || | | || | || |
Db      836 TCACCTTCCCACTCATCTATACTGTTTTCCAGCTCGTCTTGCAGCAGTAATATTAGGAA 895

Qy      920 CATATCAGACGTACAAGAGGAGATTGAAGAACAAACATG 958
      || | || || || | |   | |   || || || |
Db      896 TTTATGTCACATACAGGAAATGTTATGGAAAAATGATG 934

```

RESULT 14

AB002693

LOCUS AB002693 1629 bp mRNA linear ROD 11-AUG-1999

DEFINITION Mus musculus mRNA for ISBT, complete cds.

ACCESSION AB002693

VERSION AB002693.1 GI:1944178

KEYWORDS ISBT.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1629)
AUTHORS Saeki,T., Matoba,K., Furukawa,H., Kirifuji,K., Kanamoto,R. and
Iwami,K.
TITLE Characterization, cDNA cloning, and functional expression of mouse
ileal sodium-dependent bile acid transporter
JOURNAL J. Biochem. 125 (4), 846-851 (1999)
MEDLINE 99203592
PUBMED 10101301
REFERENCE 2 (bases 1 to 1629)
AUTHORS Saeki,T.
TITLE Direct Submission
JOURNAL Submitted (07-APR-1997) Tohru Saeki, Kyoto Prefectural University,
Department of Biological Resource Chemistry; Nakaragi, Shimogamo,
Sakyo-ku, Kyoto, Kyoto 606, Japan (E-mail:tsaeki@dns.kpu.ac.jp,
Tel:81-75-703-5663, Fax:81-75-703-5661)
FEATURES
source Location/Qualifiers
1. .1629
/organism="Mus musculus"
/mol_type="mRNA"
/strain="ICR"
/db_xref="taxon:10090"
/sex="male"
/dev_stage="8 weeks"
CDS 50. .1096
/codon_start=1
/product="ISBT"
/protein_id="BAA19606.1"
/db_xref="GI:1944179"
/translation="MDNSSVCPPNATVCEGDCSVVPESNFNAILNTVMSTVLTILLAM
VMFSMGCNVEVHKFLGHIKRPWGI FVGFLCQFGIMPLTGFI LSVASGILPVQAVVVLI
MGCCPGGTGSN I LAYWIDGDM DLSVSM T T C S T L L A L G M M P L C L F V Y T K M W V D S G T I V I
P Y D S I G I S L V A L V I P V S F G M F V N H K W P Q A K I I L K I G S I T G V I L V L I A V I G G I L Y Q S
A W I I E P K L W I I G T I F P I A G Y S L G F F L A R L A G Q P W Y R C R T V A L E T G M Q N T Q L C S T I V Q L
S F S P E D L N L V F T F P L I Y T V F Q L V F A A V I L G I Y V T Y R K C Y G K N D A E F L E K T D N E M D S R P
S F D E T N K G F Q P D E K "

```

Query Match 27.1%; Score 307.8; DB 10; Length 1629;
Best Local Similarity 59.4%; Pred. No. 1.5e-82;
Matches 522; Conservative 0; Mismatches 357; Indels 0; Gaps 0;

Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

REFERENCE 1 (bases 1 to 1116)

AUTHORS Stengelin,S., Apel,S., Becker,W., Maier,M., Rosenberger,J., Wess,G. and Kramer,W.

TITLE Cloning of the rabbit ileal sodium-dependent bile acid transporter

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1116)

AUTHORS Stengelin,S.

TITLE Direct Submission

JOURNAL Submitted (11-OCT-1995) Stengelin S., Hoechst Marion Roussel, TD Metabolism, Building H825, D-65926 Frankfurt am Main, Germany

FEATURES Location/Qualifiers

source 1. .1116

/organism="Oryctolagus cuniculus"

/mol_type="mRNA"

/strain="New Zealand White"

/db_xref="taxon:9986"

/tissue_type="ileum"

/dev_stage="adult"

5'UTR 1. .41

CDS 42. .1085

/codon_start=1

/product="ileal sodium-dependent bile acid transporter"

/protein_id="CAA91184.1"

/db_xref="GI:1019396"

/db_xref="GOA:Q28727"

/db_xref="SWISS-PROT:Q28727"

/translation="MSNLTVGCLANATVCEGASCVAPESNFNAILSVVLTSTVLTILLA
LVMFSMGCNVEIKKFLGHIRRPWGI FIGFLCQFGIMPLTGFVLAVAFGIMPIQAVVVL
IMGCCPGGTASNILAYWVDGMDLSVSMTTCTLLALGMMPLCLYVYTKMWVDSGTIV
IPYDNIGTSLVALVVPVSGMFVNHKWPQKAKIILKVGSIAGAVLIVLIAVVGILYQ
SAWIIIEPKLWIIIGTIFPMAGYSLGFFLARIAGQPWYRCRTVALETGMQNTQLCSTIVQ
LSFSPEDLTYVFTFPLIYSIFQIAFAAIFLGIYVAYRKCHGKNDAEFDPIDKDKTEPE
SSFHQMNNGGFQPE"

3'UTR 1086. .1116

ORIGIN

Query Match 27.0%; Score 306.4; DB 4; Length 1116;

Best Local Similarity 59.2%; Pred. No. 3.7e-82;

Matches 523; Conservative 0; Mismatches 361; Indels 0; Gaps 0;

Qy 80 ATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGT 139

Db 124 ATGCCATCCTCAGCGTGGTTCTGAGTACCGTGCTGACCATCCTGCTGGCTCTGGTCATGT 183

Qy 140 TCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCGCACATCAGGAGACCCTGGG 199

Db 184 TCTCCATGGGATGCAACGTGGAAATCAAGAAATTCCTGGGGCACATAAGCGGCCCTGGG 243

Qy 200 GCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGG 259

Db 244 GCATCTTCATTGGCTTCCTCTGCCAGTTTGGGATCATGCCCCCTCACGGGATTTGTCTAG 303

Qy 260 CCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCC 319

Db 304 CGGTGGCCTTTGGGATCATGCCCATCCAGGCCGTGGTGGTGTCTCATCATGGGATGCTGCC 363

Qy 320 CGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGATGGAGATATGGATCTCAGCA 379
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 364 CTGGAGGAACGGCCTCCAACATCCTGGCCTATTGGGTGGATGGAGACATGGACTTGAGTG 423

Qy 380 TCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATC 439
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 424 TCAGCATGACCACCTGCTCCACATTGCTTGCCCTCGGAATGATGCCGTTATGCCTTTATG 483

Qy 440 TCTACACCTGGTCTCTGGAGTCTTCAGCAGAATCTCACCATTCCCTTATCAGAACATAGGAA 499
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 484 TCTACACCAAAATGTGGGTGGACTCTGGGACCATTGTAATTCCTTACGATAACATAGGTA 543

Qy 500 TTACCCCTTGTGTGCCTGACCATTCTGTGGCCTTTGGTGTCTATGTGAATTACAGATGGC 559
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 544 CTTCTTGGTTGCTCTTGTGTTGTTCCCGTTTCCATCGGAATGTTTGTTAATCACAAGTGGC 603

Qy 560 CAAAACAATCCAAAATCATTTCTCAAGATTGGGGCCGTTGTTGGTGGGGTCTCCTTCTGG 619
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 604 CCCAAAAGGCGAAGATTATACTTAAAGTTGGATCCATTGCAGGTGCAGTCCTTATTGTGC 663

Qy 620 TGGTCGCAGTTGCTGGTGTGGTCTCTGGCGAAAGGATCTTGGAAATCAGACATCACCCCTTC 679
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 664 TCATAGCTGTGGTAGGAGGAATATTGTACCAAAGTGCCTGGATCATCGAACCCAAGCTGT 723

Qy 680 TGACCATCAGTTTCATCTTTCCCTTTGATTGGCCATGTCACGGGTTTTCTGCTGGCACTTT 739
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 724 GGATTATAGGGACGATATTTCCCATGGCCGGTTACTCCCTTGGCTTTTTTTTGGCTAGGA 783

Qy 740 TTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCCCTTAGAAACTGGAGCTCAGAATA 799
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 784 TAGCTGGTCAGCCATGGTACAGGTGCCGAACAGTTGCTTTGGAAACAGGGATGCAGAACA 843

Qy 800 TTCAGATGTGCATCACCATGCTCCAGTTATCTTTCACTGCTGAGCACTTGGTCCAGATGT 859
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 844 CACAGCTGTGCTCCACGATCGTGCAGCTCTCCTTCAGCCCCGAGGACCTCACTTATGTGT 903

Qy 860 TGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGATGGATTTCTTATTGTTGCAG 919
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 904 TCACCTTCCCGCTCATTTACAGCATCTTCCAGATCGCCTTTGCAGCAATCTTCTTAGGAA 963

Qy 920 CATATCAGACGTACAAGAGGAGATTGAAGAACAACATGGAAAA 963
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 964 TATATGTCGCATATAGGAAATGTCATGGGAAAAATGATGCAGAA 1007

Search completed: March 25, 2004, 17:56:19
 Job time : 4669 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2004, 08:03:39 ; Search time 575 Seconds
(without alignments)
8378.185 Million cell updates/sec

Title: US-10-091-628-1
Perfect score: 1134
Sequence: 1 atgagagccaattgttccag.....acatcacttcatgtgaatag 1134

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result		%	Query				
No.	Score	Match	Length	DB	ID	Description	
1	1134	100.0	1134	6	AAD46333	Aad46333 Human sod	
2	1134	100.0	1600	6	AAD46334	Aad46334 Human sod	
3	1130.8	99.7	1517	7	AAD47353	Aad47353 Human tra	
4	655.8	57.8	987	6	ABS59328	Abs59328 Human ile	
5	320.4	28.3	2263	2	AAQ91108	Aaq91108 Hamster i	
6	297.8	26.3	1047	2	AAQ91109	Aaq91109 Human ile	
7	297.8	26.3	3779	7	ACF63388	Acf63388 Human IBA	

	8	297.8	26.3	3779	7	ABZ20750	Abz20750 Human ile
	9	293.2	25.9	4269	9	ADB58285	Adb58285 Toxicity-
	10	293.2	25.9	4269	9	ADB52825	Adb52825 Primary r
	11	183.2	16.2	1663	6	ABK63719	Abk63719 Rat seque
	12	183.2	16.2	1663	9	ADB58234	Adb58234 Toxicity-
	13	173.6	15.3	1580	6	ABN95678	Abn95678 Gene #217
	14	173.6	15.3	1580	7	AAD56518	Aad56518 Human sod
	15	173.6	15.3	1580	7	ACC51213	Acc51213 Human Plk
	16	118.8	10.5	1413	5	AAS64762	Aas64762 DNA encod
	17	118.8	10.5	2020	7	ABZ20751	Abz20751 Human ile
	18	96.4	8.5	1368	7	ADA55865	Ada55865 Gene enco
	19	96.4	8.5	1368	7	ADA39674	Ada39674 Human sec
	20	96.4	8.5	1368	9	ADC73413	Adc73413 Human sec
	21	93.2	8.2	1399	7	ABZ36267	Abz36267 Human sec
	22	86.2	7.6	729	6	AAD33699	Aad33699 Human sec
	23	86.2	7.6	729	7	ADA56432	Ada56432 Gene enco
	24	86.2	7.6	729	7	ADA40263	Ada40263 Human sec
	25	86.2	7.6	729	9	ADC73788	Adc73788 Human sec
c	26	79.8	7.0	360	4	ABS49929	Abs49929 Human liv
c	27	79.8	7.0	560	4	ABS37088	Abs37088 Human liv
	28	76.6	6.8	1824	5	AAS64761	Aas64761 DNA encod
	29	74	6.5	401	7	ABX37088	Abx37088 Bovine ES
	30	67.6	6.0	972	5	AAH67519	Aah67519 C glutami
c	31	67.6	6.0	349980	5	AAH68532	Aah68532 C glutami
	32	66	5.8	972	7	ACA01203	Aca01203 C. glutam
	33	64.4	5.7	1017	6	ABZ12575	Abz12575 Arabidops
	34	64.4	5.7	1272	3	AAC39644	Aac39644 Arabidops
	35	64.4	5.7	1619	3	AAC49339	Aac49339 Arabidops
	36	64.2	5.7	1431	6	ABL39796	Abl39796 Human NS
	37	62.4	5.5	1152	6	ABV99337	Abv99337 Human NOV
	38	62.4	5.5	1152	6	ABV99339	Abv99339 Human NOV
	39	62.4	5.5	1317	6	AAD46335	Aad46335 Human sod
	40	62.4	5.5	1355	6	ABV99338	Abv99338 Human NOV
	41	62.4	5.5	1661	7	AAD47355	Aad47355 Human tra
	42	62.4	5.5	1777	6	AAD46336	Aad46336 Human sod
c	43	58.2	5.1	269223	4	AAF28554	Aaf28554 Genomic f
	44	57.6	5.1	2141	6	AAD22002	Aad22002 Human tra
	45	56	4.9	1870	6	ABQ61206	Abq61206 clone IMA

ALIGNMENTS

RESULT 1

AAD46333

ID AAD46333 standard; DNA; 1134 BP.

XX

AC AAD46333;

XX

DT 27-JAN-2003 (first entry)

XX

DE Human sodium/bile-like transporter DNA #1.

XX

KW Human; sodium/bile-like transporter; novel human protein; drug screening;

KW NHP; cancer; cosmetic; nutraceutical; gene therapy; cytostatic; gene;

KW chromosome 4; ds.

XX

OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1. .1134
 FT /*tag= a
 FT /product= "Human sodium/bile-like transporter"
 XX
 EN WO200272774-A2.
 XX
 PD 19-SEP-2002.
 XX
 PF 06-MAR-2002; 2002WO-US007438.
 XX
 PR 12-MAR-2001; 2001US-0275009P.
 PR 17-APR-2001; 2001US-0284152P.
 XX
 PA (LEXI-) LEXICON GENETICS INC.
 XX
 PI Wilganowski NL, Nepomnichy B, Burnett MB, Hu Y;
 XX
 DR WPI; 2002-723334/78.
 DR P-PSDB; AAE28936.
 XX
 PT New protein and nucleic acid molecule, useful for diagnosing or treating
 PT diseases, e.g. cancer, for drug screening, clinical trial monitoring,
 PT pharmacogenomics, and for cosmetic or nutraceutical applications.
 XX
 PS Claim 1; Page 37; 41pp; English.
 XX
 CC The invention relates to novel human proteins (NHP), sodium/bile-like
 CC transporter and their nucleic acids. The invention is useful for
 CC identifying the protein which may be used for diagnosis, clinical trial
 CC monitoring, drug screening, pharmacogenomics, treatment of diseases such
 CC as cancer, and for cosmetic or nutraceutical applications. The nucleic
 CC acid molecule may also be used as hybridisation probes for screening
 CC libraries, assessing gene expression patterns, and in amplification
 CC assays. It is also used in gene therapy. The present sequence is human
 CC sodium/bile-like transporter DNA. This gene is located at chromosome 4
 XX
 SQ Sequence 1134 BP; 251 A; 281 C; 287 G; 315 T; 0 U; 0 Other;

Query Match 100.0%; Score 1134; DB 6; Length 1134;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG	60
Db	1	ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG	60
Qy	61	CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTG	120
Db	61	CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTG	120
Qy	121	ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTTCG	180
Db	121	ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTTCG	180

Qy 181 CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT 240
 |||
 Db 181 CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT 240

Qy 241 TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT 300
 |||
 Db 241 TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT 300

Qy 301 CTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGAT 360
 |||
 Db 301 CTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGAT 360

Qy 361 GGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATG 420
 |||
 Db 361 GGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATG 420

Qy 421 ATGCCACTCTGCATTTATCTCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATT 480
 |||
 Db 421 ATGCCACTCTGCATTTATCTCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATT 480

Qy 481 CCTTATCAGAACATAGGAATTACCCTTGTGTGCCTGACCATTCCCTGTGGCCTTTGGTGTC 540
 |||
 Db 481 CCTTATCAGAACATAGGAATTACCCTTGTGTGCCTGACCATTCCCTGTGGCCTTTGGTGTC 540

Qy 541 TATGTGAATTACAGATGGCCAAAACAATCCAAATCATTCTCAAGATTGGGGCCGTTGTT 600
 |||
 Db 541 TATGTGAATTACAGATGGCCAAAACAATCCAAATCATTCTCAAGATTGGGGCCGTTGTT 600

Qy 601 GGTGGGGTCCTCCTTCTGGTGGTGCAGTTGCTGGTGTGGTCCCTGGCGAAAGGATCTTGG 660
 |||
 Db 601 GGTGGGGTCCTCCTTCTGGTGGTGCAGTTGCTGGTGTGGTCCCTGGCGAAAGGATCTTGG 660

Qy 661 AATTCAGACATCACCCCTTCTGACCATCAGTTTCATCTTTCCTTTGATTGGCCATGTCACG 720
 |||
 Db 661 AATTCAGACATCACCCCTTCTGACCATCAGTTTCATCTTTCCTTTGATTGGCCATGTCACG 720

Qy 721 GGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTCCTTA 780
 |||
 Db 721 GGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTCCTTA 780

Qy 781 GAAACTGGAGCTCAGAATATTCAGATGTGCATCACCATGCTCCAGTTATCTTCACTGCT 840
 |||
 Db 781 GAAACTGGAGCTCAGAATATTCAGATGTGCATCACCATGCTCCAGTTATCTTCACTGCT 840

Qy 841 GAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGAT 900
 |||
 Db 841 GAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGAT 900

Qy 901 GGATTTCTTATTGTTGCAGCATATCAGACGTACAAGAGGAGATTGAAGAACAAACATGGA 960
 |||
 Db 901 GGATTTCTTATTGTTGCAGCATATCAGACGTACAAGAGGAGATTGAAGAACAAACATGGA 960

Qy 961 AAAAAGAACTCAGGTTGCACAGAAGTCTGCCATACGAGGAAATCGACTTCTTCCAGAGAG 1020
 |||
 Db 961 AAAAAGAACTCAGGTTGCACAGAAGTCTGCCATACGAGGAAATCGACTTCTTCCAGAGAG 1020

Qy 1021 ACCAATGCCTTCTTGGAGGTGAATGAAGAAGGTGCCATCACTCCTGGGCCACCAGGGCCA 1080

```

Db      1021  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          ACCAATGCCTTCTTGGAGGTGAATGAAGAAGGTGCCATCACTCCTGGGCCACCAGGGCCA 1080

Qy      1081  ATGGATTGCCACAGGGCTCTCGAGCCAGTTGGCCACATCACTTCATGTGAATAG 1134
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      1081  ATGGATTGCCACAGGGCTCTCGAGCCAGTTGGCCACATCACTTCATGTGAATAG 1134

```

RESULT 2

AAD46334

ID AAD46334 standard; DNA; 1600 BP.

XX

AC AAD46334;

XX

DT 27-JAN-2003 (first entry)

XX

DE Human sodium/bile-like transporter DNA #2.

XX

KW Human; sodium/bile-like transporter; novel human protein; drug screening;

KW NHP; cancer; cosmetic; nutraceutical; gene therapy; cytostatic; gene;

KW chromosome 4; ds.

XX

OS Homo sapiens.

XX

PN WO200272774-A2.

XX

PD 19-SEP-2002.

XX

PF 06-MAR-2002; 2002WO-US007438.

XX

PR 12-MAR-2001; 2001US-0275009P.

PR 17-APR-2001; 2001US-0284152P.

XX

PA (LEXI-) LEXICON GENETICS INC.

XX

PI Wilganowski NL, Nepomnichy B, Burnett MB, Hu Y;

XX

DR WPI; 2002-723334/78.

XX

PT New protein and nucleic acid molecule, useful for diagnosing or treating

PT diseases, e.g. cancer, for drug screening, clinical trial monitoring,

PT pharmacogenomics, and for cosmetic or nutraceutical applications.

XX

PS Disclosure; Page 38-39; 41pp; English.

XX

CC The invention relates to novel human proteins (NHP), sodium/bile-like

CC transporter and their nucleic acids. The invention is useful for

CC identifying the protein which may be used for diagnosis, clinical trial

CC monitoring, drug screening, pharmacogenomics, treatment of diseases such

CC as cancer, and for cosmetic or nutraceutical applications. The nucleic

CC acid molecule may also be used as hybridisation probes for screening

CC libraries, assessing gene expression patterns, and in amplification

CC assays. It is also used in gene therapy. The present sequence is human

CC sodium/bile-like transporter DNA. This gene is located at chromosome 4

XX

SQ Sequence 1600 BP; 367 A; 366 C; 399 G; 468 T; 0 U; 0 Other;

Query Match 100.0%; Score 1134; DB 6; Length 1600;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG 60
        |||
Db     194 ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG 253

Qy      61 CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTACAGTGGTGTCCACTGTG 120
        |||
Db     254 CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTACAGTGGTGTCCACTGTG 313

Qy     121 ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCG 180
        |||
Db     314 ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCG 373

Qy     181 CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT 240
        |||
Db     374 CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT 433

Qy     241 TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT 300
        |||
Db     434 TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT 493

Qy     301 CTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGAT 360
        |||
Db     494 CTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGAT 553

Qy     361 GGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATG 420
        |||
Db     554 GGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATG 613.

Qy     421 ATGCCACTCTGCATTTATCTCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATT 480
        |||
Db     614 ATGCCACTCTGCATTTATCTCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATT 673

Qy     481 CCTTATCAGAACATAGGAATTACCCTTGTGTGCCTGACCATTCCCTGTGGCCTTTGGTGTC 540
        |||
Db     674 CCTTATCAGAACATAGGAATTACCCTTGTGTGCCTGACCATTCCCTGTGGCCTTTGGTGTC 733

Qy     541 TATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTT 600
        |||
Db     734 TATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTT 793

Qy     601 GGTGGGGTCCTCCTTCTGGTGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGATCTTGG 660
        |||
Db     794 GGTGGGGTCCTCCTTCTGGTGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGATCTTGG 853

Qy     661 AATTCAGACATCACCTTCTGACCATCAGTTTCATCTTTCCTTTGATTGGCCATGTCACG 720
        |||
Db     854 AATTCAGACATCACCTTCTGACCATCAGTTTCATCTTTCCTTTGATTGGCCATGTCACG 913

Qy     721 GGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCTTA 780
        |||
Db     914 GGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCTTA 973

Qy     781 GAAACTGGAGCTCAGAATATTCAGATGTGCATCACCATGCTCCAGTTATCTTTCACTGCT 840
```

```

Db      974  |||||GAAACTGGAGCTCAGAATATTCAGATGTGCATCACCATGCTCCAGTTATCTTTCAGTGCT 1033
Qy      841  GAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGAT 900
Db      1034  |||||GAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGAT 1093
Qy      901  GGATTTCTTATTGTTGCAGCATATCAGACGTACAAGAGGAGATTGAAGAACAAACATGGA 960
Db      1094  |||||GGATTTCTTATTGTTGCAGCATATCAGACGTACAAGAGGAGATTGAAGAACAAACATGGA 1153
Qy      961  AAAAAAGAACTCAGGTTGCACAGAAGTCTGCCATACGAGGAAATCGACTTCTTCCAGAGAG 1020
Db      1154  |||||AAAAAGAACTCAGGTTGCACAGAAGTCTGCCATACGAGGAAATCGACTTCTTCCAGAGAG 1213
Qy      1021  ACCAATGCCTTCTTGGAGGTGAATGAAGAAGGTGCCATCACTCCTGGGCCACCAGGGCCA 1080
Db      1214  |||||ACCAATGCCTTCTTGGAGGTGAATGAAGAAGGTGCCATCACTCCTGGGCCACCAGGGCCA 1273
Qy      1081  ATGGATTGCCACAGGGCTCTCGAGCCAGTTGGCCACATCACTTCATGTGAATAG 1134
Db      1274  |||||ATGGATTGCCACAGGGCTCTCGAGCCAGTTGGCCACATCACTTCATGTGAATAG 1327

```

RESULT 3

AAD47353

ID AAD47353 standard; cDNA; 1517 BP.

XX

AC AAD47353;

XX

DT 24-FEB-2003 (first entry)

XX

DE Human transporter and ion channel (TRICH) cDNA #6.

XX

KW Human; transporter and ion channel; TRICH; neurodegenerative disorder;
KW Parkinson's disease; Alzheimer's disease; muscular disorder; transgenic;
KW myotonic dystrophy; catatonia; endocrine disorder; diabetes; cytostatic;
KW Grave's disease; cancer; leukaemia; cervical; immunological; scleroderma;
KW systemic lupus erythematosus; allergy; gastrointestinal; Crohn's disease;
KW Goodpasture's syndrome; infection; cardiovascular; fungicide; nootropic;
KW hepatic disease; cirrhosis; gene therapy; uropathic; anti-HIV; virucide;
KW atherosclerosis; antiparasitic; protozoacide; antibacterial; gene; ss.

XX

OS Homo sapiens.

XX

FH	Key	Location/Qualifiers
FT	CDS	249..1382
FT		/*tag= a
FT		/product= "Human TRICH protein"
FT	sig_peptide	361..539
FT		/*tag= b
FT	mat_peptide	540..1379
FT		/*tag= c
FT		/product= "Human mature TRICH protein"

XX

PN WO200277237-A2.

XX

PD 03-OCT-2002.
 XX
 PF 08-FEB-2002; 2002WO-US003657.
 XX
 PR 09-FEB-2001; 2001US-0267892P.
 PR 23-FEB-2001; 2001US-0271168P.
 PR 02-MAR-2001; 2001US-0272890P.
 PR 16-MAR-2001; 2001US-0276860P.
 PR 23-MAR-2001; 2001US-0278255P.
 PR 30-MAR-2001; 2001US-0280538P.
 PR 25-JAN-2002; 2002US-0351359P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Lee EA, Ding L, Baughn MR, Tribouley CM, Bruns CM, Elliott VS;
 PI Walia NK, Forsythe IJ, Raumann BE, Burford N, Lal PG, Thornton M;
 PI Gandhi AR, Arvizu C, Yao MG, Yue H, Xu Y, Hafalia AJA, Ison CH;
 PI Chen H;
 XX
 DR WPI; 2003-018931/01.
 DR P-PSDB; AAE29906.
 XX
 PT New TRICH polypeptides, useful for diagnosing, preventing, and treating
 PT disorders associated with an abnormal expression or activity of TRICH,
 PT e.g. neuromuscular, immunological, cardiovascular disorders, cancer and
 PT infection.
 XX
 PS Claim 5; Page 197-198; 214pp; English.
 XX
 CC The invention relates to human transporters and ion channels (TRICH) and
 CC their nucleic acids. The sequences of the invention are useful in
 CC diagnosing, preventing, and treating disorders associated with an
 CC abnormal expression or activity of TRICH, such as neurodegenerative
 CC disorders (e.g. Parkinson's disease, Alzheimer's disease), muscular
 CC disorders (e.g. myotonic dystrophy, catatonia), endocrine disorders (e.g.
 CC diabetes, Grave's disease), cancers (e.g. leukaemia, cervical or breast
 CC cancers), immunological disorders (e.g. scleroderma, systemic lupus
 CC erythematosus, allergies), gastrointestinal disorders (e.g. Crohn's
 CC disease), renal disorders (e.g. Goodpasture's syndrome), infections (e.g.
 CC viral, bacterial, fungal, parasitic, protozoal, helminthic),
 CC cardiovascular disorders (e.g. atherosclerosis), or hepatic diseases
 CC (e.g. cirrhosis). TRICH or its fragments may also be used in screening
 CC for compounds that specifically bind to and modulate its activity. TRICH
 CC DNA can be used to create humanised animals or transgenic animals to
 CC model human disease. It is also used in gene therapy. The present
 CC sequence is human TRICH cDNA
 XX
 SQ Sequence 1517 BP; 356 A; 352 C; 374 G; 435 T; 0 U; 0 Other;

 Query Match 99.7%; Score 1130.8; DB 7; Length 1517;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1132; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

 Qy 1 ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 249 ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG 308

Qy 61 CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTG 120
 |||
 Db 309 CCAGTGGGACTGGAGGCGCATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCCACTGTG 368

Qy 121 ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCTG 180
 |||
 Db 369 ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCTG 428

Qy 181 CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT 240
 |||
 Db 429 CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT 488

Qy 241 TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT 300
 |||
 Db 489 TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT 548

Qy 301 CTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTGAT 360
 |||
 Db 549 CTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTGAT 608

Qy 361 GGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATG 420
 |||
 Db 609 GGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATG 668

Qy 421 ATGCCACTCTGCATTTATCTCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATT 480
 |||
 Db 669 ATGCCACTCTGCATTTATCTCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATT 728

Qy 481 CCTTATCAGAACATAGGAATTACCCTTGTGTGCCTGACCATTCTGTGGCCTTTGGTGTG 540
 |||
 Db 729 CCTTATCAGAACATAGGAATTACCCTTGTGTGCCTGACCATTCTGTGGCCTTTGGTGTG 788

Qy 541 TATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTT 600
 |||
 Db 789 TATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTT 848

Qy 601 GGTGGGGTCCTCCTTCTGGTGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGATCTTGG 660
 |||
 Db 849 GGTGGGGTCCTCCTTCTGGTGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGATCTTGG 908

Qy 661 AATTCAGACATCACCTTCTGACCATCAGTTTCATCTTTCCTTTGATTGGCCATGTCACG 720
 |||
 Db 909 AATTCAGACATCACCTTCTGACCATCAGTTTCATCTTTCCTTTGATTGGCCATGTCACG 968

Qy 721 GGTTTTCTGCTGGCACTTTTTTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTCCTTA 780
 |||
 Db 969 GGTTTTCTGCTGGCACTTTTTTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTCCTTA 1028

Qy 781 GAAACTGGAGCTCAGAATATTAGATGTGCATCACCATGCTCCAGTTATCTTTCAGTGCT 840
 |||
 Db 1029 GAAACTGGAGCTCAGAATATTAGATGTGCATCACCATGCTCCAGTTATCTTTCAGTGCT 1088

Qy 841 GAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGAT 900
 |||
 Db 1089 GAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGAT 1148

Qy 901 GGATTTCTTATTGTTGCAGCATATCAGACGTACAAGAGGAGATTGAAGAACAAACATGGA 960

Db	1149		GGATTTCTTATTGTTGCAGCATATCAGACGTACAAGAGGAGATTGAAGAACAAACATGGA	1208
Qy	961		AAAAAGAACTCAGGTTGCACAGAAGTCTGCCATACGAGGAAATCGACTTCTTCCAGAGAG	1020
Db	1209		AAAAAGAACTCAGGTTGCACAGAAGTCTGCCATACGAGGAAATCGACTTCTTCCAGAGAG	1268
Qy	1021		ACCAATGCCTTCTTGGAGGTGAATGAAGAAGGTGCCATCACTCCTGGGCCACCAGGGCCA	1080
Db	1269		ACCAATGCCTTCTTGGAGGTGAATGAAGAAGGTGCCATCACTCCTGGGCCACCAGGGCCA	1328
Qy	1081		ATGGATTGCCACAGGGCTCTCGAGCCAGTTGGCCACATCACTTCATGTGAATAG	1134
Db	1329		ATGGATTGCCACAGGGCTCTCGAGCCAGTTGGCCACATCACTTCATGTGAATAG	1382

RESULT 4

ABS59328

ID ABS59328 standard; DNA; 987 BP.

XX

AC ABS59328;

XX

DT 05-NOV-2002 (first entry)

XX

DE Human ileal sodium/bile acid cotransporter-like gene.

XX

KW Human; NOX; cardiomyopathy; atherosclerosis; cell signal processing;
KW breast cancer; Alzheimer's disease; epilepsy; Huntington's disease;
KW anxiety; behavioural disorder; multiple sclerosis; myasthenia gravis;
KW neurodegeneration; Parkinson's disease; pain; stroke; endometriosis;
KW autoimmune disease; allergy; addiction; asthma; transplantation;
KW graft versus host disease; systemic lupus erythematosus; scleroderma;
KW psoriasis; Crohn's disease; HIV infection; human immunodeficiency virus;
KW atherosclerosis; cirrhosis; rheumatoid arthritis; diabetes; pancreatitis;
KW thrombocytopenia; bleeding disorder; metabolic disorder; obesity;
KW glucose transport defect; glomerulonephritis; hypercalcaemia;
KW polycystic kidney disease; renal tubular acidosis; skin disorder;
KW congenital diarrhoea; respiratory disease; gastro-intestinal disease;
KW muscle disorder; bone disorder; joint disorder; skeletal disorder;
KW haematopoietic disorder; urinary system disorder; osteoporosis; ds;
KW dental disease; dental infection; growth disorder; reproductive disorder;
KW hypogonadism; fertility disorder; viral infection; bacterial infection;
KW parasitic infection; metabolic pathway modulation; gene therapy; gene;
KW zinc metalloprotease; ADAM-TS 7; alpha-2-macroglobulin precursor;
KW ileal sodium/bile acid cotransporter; prohibitin; MT; CIP4; spinesin;
KW macrophage stimulating protein precursor; fatty acid-binding protein;
KW gap junction beta-5 protein; hepsin/plasma transmembrane serine protease.

XX

OS Homo sapiens.

XX

PN WO200233087-A2.

XX

PD 25-APR-2002.

XX

PF 17-OCT-2001; 2001WO-US032496.

XX

PR 17-OCT-2000; 2000US-0241040P.

PR 17-OCT-2000; 2000US-0241058P.
PR 17-OCT-2000; 2000US-0241063P.
PR 17-OCT-2000; 2000US-0241243P.
PR 20-OCT-2000; 2000US-0242152P.
PR 23-OCT-2000; 2000US-0242482P.
PR 23-OCT-2000; 2000US-0242611P.
PR 23-OCT-2000; 2000US-0242612P.
PR 24-OCT-2000; 2000US-0242880P.
PR 24-OCT-2000; 2000US-0242881P.
PR 29-DEC-2000; 2000US-0259028P.
PR 20-FEB-2001; 2001US-0269813P.
PR 25-APR-2001; 2001US-0286324P.
PR 29-MAY-2001; 2001US-0294108P.
PR 09-JUL-2001; 2001US-0303698P.
PR 16-OCT-2001; 2001US-00981151.

XX

PA (CURA-) CURAGEN CORP.

XX

PI Edinger S, Gerlach V, Macdougall JR, Malyankar UM, Smithson G;
PI Millet I, Peyman JA, Stone DJ, Gunther E, Ellerman K, Shimkets RA;
PI Padigar M, Guo X, Patturajan M, Taupier RJ, Burgess CE;
PI Zerhusen BD, Kekuda R, Spytek KA, Gangolli EA, Fernandes ER;
PI Gorman L;

XX

DR WPI; 2002-590434/63.

DR P-PSDB; ABG76899.

XX

PT Cytoplasmic, nuclear, membrane bound and secreted polypeptides and
PT nucleic acids encoding the polypeptides for diagnosing and treating e.g.
PT cancer, Alzheimer's disease, cardiomyopathy, metabolic disease and
PT diabetes.

XX

PS Claim 8; Page 50; 305pp; English.

XX

CC The present invention relates to new NOVX (NOV1-10) polypeptides. The
CC molecules of the invention are useful for treating or preventing a NOVX-
CC associated disorder, such as cardiomyopathy, atherosclerosis, or a
CC disorder related to cell signal processing and metabolic pathway
CC modulation in humans. NOVX polypeptides, nucleic acids and antibodies are
CC useful for treating or preventing disorders or syndromes including breast
CC cancer, Alzheimer's disease, epilepsy, Huntington's disease, anxiety,
CC behavioural disorders, multiple sclerosis, myasthenia gravis,
CC neurodegeneration, Parkinson's disease, pain, stroke, autoimmune disease,
CC allergies, addiction, asthma, endometriosis, graft versus host disease,
CC systemic lupus erythematosus, scleroderma, transplantation, psoriasis,
CC Crohn's disease, HIV (human immunodeficiency virus) infection,
CC atherosclerosis, cirrhosis, rheumatoid arthritis, diabetes,
CC thrombocytopenia, bleeding disorders, metabolic disorders, obesity,
CC glucose transport defect, glomerulonephritis, hypercalcaemia, polycystic
CC kidney disease, pancreatitis, renal tubular acidosis, skin disorders,
CC congenital diarrhoea, respiratory disease, gastro-intestinal diseases,
CC muscle, bone, joint and skeletal disorders, haematopoietic disorders,
CC urinary system disorders, osteoporosis, dental disease and infection,
CC growth and reproductive disorders, hypogonadism, fertility, and/or other
CC pathologies and disorders, viral, bacterial, or parasitic infections. The
CC present nucleic acid sequence encodes a NOVX protein of the invention

XX

SQ Sequence 987 BP; 206 A; 243 C; 236 G; 302 T; 0 U; 0 Other;

Query Match 57.8%; Score 655.8; DB 6; Length 987;
Best Local Similarity 86.3%; Pred. No. 1.5e-195;
Matches 803; Conservative 0; Mismatches 77; Indels 51; Gaps 5;

```
Qy      1 ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG 60
        |||
Db      1 ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG 60

Qy     61 CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTCACAGTGGTGTCCACTGTG 120
        |||
Db     61 CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTCACAGTGGTGTCCACTATC 120

Qy    121 ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTTCG 180
        |||
Db    121 ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTTCG 180

Qy    181 CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT 240
        |||
Db    181 CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT 240

Qy    241 TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT 300
        |||
Db    241 TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT 300

Qy    301 CTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGAT 360
        |||
Db    301 CTCATCATGGGCTGCTG-CCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGAT 359

Qy    361 GGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATG 420
        |||
Db    360 GGAGATATGGATCTCA-----GGTGCCCTGGGAATG 390

Qy    421 ATGCCACTCTGCATTTATCTCTACACCTGGTCTGGAGTCTTCAGCAGAAATCTCACCATT 480
        |||
Db    391 ATGCCACTCTGCATTTATCTCTACACCTGGTCTGGAGTCTTCAGCAGAAATCTCACCATT 450

Qy    481 CCTTATCAGAAC-----TAGGAATTACCCTTGTGTGCCTGACCATTCCTGTG 528
        |||
Db    451 CCTTATCAGAACATAGGTCTGTCTTTAGGAATTACCCTTGTGTGCCTGACCATTCCTGTG 510

Qy    529 GCCTTTGGTGTCTATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTTCTCAAGATT 588
        |||
Db    511 GCCTTTGGTGTCTATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTTCTCAA---- 566

Qy    589 GGGGCCGTTGTTGGTGGGGTCCTCCTTCTGGTGGTGCAGTTGCTGGTGTGGTCTGGCG 648
        |||
Db    567 --GGCCGTTGTTGGTGGGGTCCTCCTTCTGGTGGTGCAGTTGCTGGTGTGGTCTGGCG 624

Qy    649 AAAGGATCTTGAATTCAGACATCACCCCTTCTGACCATCAGTTTCATCTTCTCTTGTGATT 708
        |||
Db    625 AAAGGATCTTGAATTCAGACATCACCCCTTCTGACCATCAGTTTCATCTTCTCTTGTGATT 684

Qy    709 GGCCATGTCACGGGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGCCAAAGGTGCAGG 768
        |||
Db    685 GGCCATGTCACGGGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGCCAAAGGACCTTG 744
```

Qy 769 ACAATTTCTTAGAACTGGAGCTCAGAATATTCAGATGTGCATCACCATGCTCCAGTTA 828
 | | | | | | | | | | | | | | | | | |
 Db 745 CCTATCTTTTGTAG---GTTTAGCTTCAAGACACCCCTGTGATACCCTACTCGCAATGACT 801
 Qy 829 TCTTTTCACTGCTGAGCATTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTC 888
 | | | | | | | | | | | | | | | | | |
 Db 802 TCGTGTCTGAATGTTCCAGGCTCATCTATGCCTTCATTCTCTGCTATATGGACTCTTC 861
 Qy 889 CAGCTGATAGATGGATTTCTTATTGTTGCAG 919
 | | | | | | | | | | | | | | | | | |
 Db 862 CAGCTGATAGATGGATTTCTTATTGTTGAAG 892

RESULT 5

AAQ91108

ID AAQ91108 standard; cDNA; 2263 BP.

XX

AC AAQ91108;

XX

DT 17-DEC-1995 (first entry)

XX

DE Hamster ileal/renal bile acid cotransporter.

XX

KW Ileal/renal bile acid cotransporter; therapeutic; gene therapy;

KW diagnostic; ss.

XX

OS Cricetulus griseus.

XX

FH Key Location/Qualifiers

FT CDS 109. .1152

FT /*tag= a

XX

PN W09517905-A1.

XX

PD 06-JUL-1995.

XX

PF 29-DEC-1994; 94WO-US014431.

XX

PR 29-DEC-1993; 93US-00176126.

XX

PA (UYWA-) UNIV WAKE FOREST.

XX

PI Dawson PA;

XX

DR WPI; 1995-246189/32.

DR P-PSDB; AAR77224.

XX

PT Hamster and human ileal and bile acid transport DNA and protein - useful
 PT in treatment of e.g. hypercholesterolaemia, diabetes and various
 PT digestive diseases, and in gene therapy to restore bile acid uptake
 PT activity.

XX

PS Claim 4; Page 98-103; 148pp; English.

XX

CC The ileal/renal bile acid cotransporter cDNA is cloned in an expression
 CC vector (plasmid pCMX or plasmid pCMV5) under the control of a baculo

CC virus Autographa californica nuclear-polyhedrosis virus gene promoter,
 CC the cytomegalo virus immediate early gene promoter, the SV40 virus late
 CC gene promoter or an inducible promoter e.g. the lactose operon promoter,
 CC and expressed in CHO, MDCK, CaCo2, BHK or preferably COS-1A cells. The
 CC cotransporter is useful in the treatment of hypercholesterolaemia,
 CC diabetes, heart disease, liver disease and various digestive disorders.
 CC The cDNA may be used in gene therapy to restore bile acid uptake activity
 CC to patients whose ileum has been surgically resected for diseases such as
 CC Crohn disease, patients born with congenital defects in the bile
 CC transporter, and patients suffering from adult-onset chronic idiopathic
 CC bile acid diarrhoea. The DNA and protein may be used in screening methods
 CC as modulators of ileal/renal bile acid cotransport activity. The DNA can
 CC also be used to detect mutations and RFLPs in human ileal/renal bile acid
 CC cotransporter genes by amplification with primers (see AAQ91110-15)

XX

SQ Sequence 2263 BP; 672 A; 451 C; 476 G; 664 T; 0 U; 0 Other;

Query Match 28.3%; Score 320.4; DB 2; Length 2263;
 Best Local Similarity 60.8%; Pred. No. 1.3e-89;
 Matches 522; Conservative 0; Mismatches 336; Indels 0; Gaps 0;

Qy	80	ATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGT	139
Db	188	ACGCCATCCTCAGCGTGGTGATGAGCACCGTGCTCACAATCCTCCTAGCCTTGGTGATGT	247
Qy	140	TCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCGCACATCAGGAGACCCTGGG	199
Db	248	TTCCATGGGGTGCAATGTGGAATCCACAAGTTTCTGGGACACCTAAGGCGCCATGGG	307
Qy	200	GCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGG	259
Db	308	GCATCGTCGTGGGCTTCCTCTGTCTGAGTTTGAATCATGCCTCTCACAGGTTTCGTCTGT	367
Qy	260	CCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCC	319
Db	368	CCGTGGCCTTTGGCATCCTCCAGTGCAAGCTGTGGTGGTGCTGATCCAGGGTTGCTGCC	427
Qy	320	CGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGATGGAGATATGGATCTCAGCA	379
Db	428	CTGGAGGAACTGCCTCCAATATCCTAGCCTATTGGGTAGATGGCGACATGGACCTCAGCG	487
Qy	380	TCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATC	439
Db	488	TTAGCATGACCACCTGCTCCACGCTGCTTGCCCTTGAATGATGCCCTTTGCCTCTTCA	547
Qy	440	TCTACACCTGGTCTGGAGTCTTCAGCAGAATCTCACCATTCTTATCAGAACATAGGAA	499
Db	548	TCTATACCAAGATGTGGGTTGACTCAGGGACGATTGTGATTCTTATGACAGCATTGGCA	607
Qy	500	TTACCCTTGTGTGCCTGACCATTCTGTGGCCTTTGGTGTCTATGTGAATTACAGATGGC	559
Db	608	CTTCTCTGGTTGCTCTTGTATTCTGTTTCCATTGGAATGTATGTGAATCACAAATGGC	667
Qy	560	CAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTTGGTGGGGTCCTCCTTCTGG	619
Db	668	CCCAAAAAGCAAAGATCATACTTAAATTTGGATCCATCGCAGGTGCAATTCTCATTGTTT	727

Qy 620 TGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGATCTTGAATTCAGACATCACCCCTTC 679
 | ||| | | | | | | | | | | | | | | | | | |
 Db 728 TCATCGCTGTGGTTGGAGGAATACTGTACCAAAGTGCCTGGACCATTGAACCCAAGCTGT 787
 Qy 680 TGACCATCAGTTTCATCTTTCCTTTGATTGGCCATGTCACGGGTTTTCTGCTGGCACTTT 739
 | | | | | | | | | | | | | | | | | | | | | |
 Db 788 GGATTATAGGAACCATATATCCTATAGCTGGCTACGGCCTGGGGTTTTCTGGCTAGAA 847
 Qy 740 TTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCCCTTAGAAACTGGAGCTCAGAATA 799
 | | | | | | | | | | | | | | | | | | | | | |
 Db 848 TTGCTGGTCAACCCTGGTACAGGTGCCGAACAGTTGCCTTGGAACCGGGTTGCAGAACA 907
 Qy 800 TTCAGATGTGCATCACCATGCTCCAGTTATCTTCACTGCTGAGCACTTGGTCCAGATGT 859
 | | | | | | | | | | | | | | | | | | | | | |
 Db 908 CTCAGCTGTGTTCCACCATTGTGCAGCTTTCCTTCAGCCCTGAGGACCTCAACCTTGTGT 967
 Qy 860 TGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGATGGATTTCTTATTGTTGCAG 919
 | | | | | | | | | | | | | | | | | | | | | |
 Db 968 TCACCTTCCCCCTCATCTACAGCATCTTCCAGATCGCCTTTGCAGCAATACTATTAGGAG 1027
 Qy 920 CATATCAGACGTACAAGA 937
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1028 CTTATGTCGCATACAAGA 1045

RESULT 6

AAQ91109

ID AAQ91109 standard; cDNA; 1047 BP.

XX

AC AAQ91109;

XX

DT 17-DEC-1995 (first entry)

XX

DE Human ileal/renal bile acid cotransporter.

XX

KW Ileal/renal bile acid cotransporter; therapeutic; gene therapy;

KW diagnostic; ss.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 1. .1044

FT /*tag= a

XX

PN WO9517905-A1.

XX

PD 06-JUL-1995.

XX

PF 29-DEC-1994; 94WO-US014431.

XX

PR 29-DEC-1993; 93US-00176126.

XX

PA (UYWA-) UNIV WAKE FOREST.

XX

PI Dawson PA;

XX

DR WPI; 1995-246189/32.

DR P-PSDB; AAR77225.

XX

PT Hamster and human ileal and bile acid transport DNA and protein - useful
PT in treatment of e.g. hypercholesterolaemia, diabetes and various
PT digestive diseases, and in gene therapy to restore bile acid uptake
PT activity.

XX

PS Claim 5; Page 107-111; 148pp; English.

XX

CC The ileal/renal bile acid cotransporter cDNA is cloned in an expression
CC vector (plasmid pCMX or plasmid pCMV5) under the control of a baculo
CC virus Autographa californica nuclear-polyhedrosis virus gene promoter,
CC the cytomegalo virus immediate early gene promoter, the SV40 virus late
CC gene promoter or an inducible promoter e.g. the lactose operon promoter,
CC and expressed in CHO, MDCK, CaCo2, BHK or preferably COS-1A cells. The
CC cotransporter is useful in the treatment of hypercholesterolaemia,
CC diabetes, heart disease, liver disease and various digestive disorders.
CC The cDNA may be used in gene therapy to restore bile acid uptake activity
CC to patients whose ileum has been surgically resected for diseases such as
CC Crohn disease, patients born with congenital defects in the bile
CC transporter, and patients suffering from adult-onset chronic idiopathic
CC bile acid diarrhoea. The DNA and protein may be used in screening methods
CC as modulators of ileal/renal bile acid cotransport activity. The DNA can
CC also be used to detect mutations and RFLPs in human ileal/renal bile acid
CC cotransporter genes by amplification with primers (see AAQ91110-15)

XX

SQ Sequence 1047 BP; 251 A; 251 C; 255 G; 290 T; 0 U; 0 Other;

Query Match 26.3%; Score 297.8; DB 2; Length 1047;
Best Local Similarity 58.5%; Pred. No. 1.1e-82;
Matches 518; Conservative 0; Mismatches 367; Indels 0; Gaps 0;

Qy	80	ATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGT	139
Db	80	ATAACATCCTAAGTGTGGTCCTAAGTACGGTGCTGACCATCCTGTTGGCCTTGGTGATGT	139
Qy	140	TCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCGCACATCAGGAGACCCTGGG	199
Db	140	TCTCCATGGGATGCAACGTGGAATCAAGAAATTTCTAGGGCACATAAAGCGGCCGTGGG	199
Qy	200	GCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGG	259
Db	200	GCATTTGTGTTGGCTTCCTCTGTTCAGTTTGGGAATCATGCCCTCACAGGATTCATCCTGT	259
Qy	260	CCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCC	319
Db	260	CGGTGGCCTTTGACATCCTCCCGCTCCAGGCCGTAGTGGTGCTCATTATAGGATGCTGCC	319
Qy	320	CGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGATGGAGATATGGATCTCAGCA	379
Db	320	CTGGAGGAAGTGCCTCCAATATCTTGGCCTATTGGGTCGATGGCGACATGGACCTGAGCG	379
Qy	380	TCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATC	439
Db	380	TCAGCATGACCACATGCTCCACACTGCTTGCCCTCGGAATGATGCCGCTGTGCCTCCTTA	439
Qy	440	TCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATTCCTTATCAGAACATAGGAA	499

```

      |||| |||      |||      ||  ||  |||| ||| | ||||| |||
Db      440 TCTATACCAAATGTGGGTCGACTCTGGGAGCATCGTAATCCCTATGATAACATAGGTA 499

Qy      500 TTACCCTTGTGTGCCTGACCATTCCTGTGGCCTTTGGTGTCTATGTGAATTACAGATGGC 559
      | ||| ||      ||      ||||| || |||| | | ||| ||| ||| ||||
Db      500 CATCTCTGGTTGCTCTCGTTGTTCCCTGTTCCATTGGAATGTTTGTTAATCACAAATGGC 559

Qy      560 CAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTTGGTGGGGTCCTCCTTCTGG 619
      | || ||| | || |||| || || |||| || | | ||| ||| |||
Db      560 CCCAAAAGCAAAGATCATACTTAAATTTGGGTCCATCGCGGGCGCCATCCTCATTGTGC 619

Qy      620 TGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGATCTTGAATTCAGACATCACCCCTTC 679
      | | || || || ||| | || || || | |||| | | || ||
Db      620 TCATAGCTGTGGTTGGAGGAATATTGTACCAAAGCGCTGGATCATTGCTCCCAAATGT 679

Qy      680 TGACCATCAGTTTCATCTTTCTTTGATTGGCCATGTCACGGGTTTTCTGCTGGCACTTT 739
      || || | | || |||| || || | | || |||| ||||
Db      680 GGATTATAGGAACAATATTTCTGTGGCGGGTTACTCCCTGGGGTTCTTCTGGCTAGAA 739

Qy      740 TTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCTTAGAAACTGGAGCTCAGAATA 799
      || | | | ||| | ||||| || || || || |||| || |||| |
Db      740 TTGCTGGTCTACCCTGGTACAGGTGCCGAACGGTTGCTTTTGAAACGGGGATGCAGAACA 799

Qy      800 TTCAGATGTGCATCACCATGCTCCAGTTATCTTTCACTGCTGAGCACTTGGTCCAGATGT 859
      ||| | || ||||| | ||| | || |||| |||| | | ||
Db      800 CGCAGCTATGTTCCACCATCGTTCAGCTCTCCTTCACTCCTGAGGAGCTCAATGTCTGAT 859

Qy      860 TGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGATGGATTCTTATTGTTGCAG 919
      | | |||| || ||| | | ||||| || || | | |||
Db      860 TCACCTTCCCGCTCATCTACAGCATTTTCCAGCTCGCCTTTGCCGCAATATTCTTAGGAT 919

Qy      920 CATATCAGACGTACAAGAGGAGATTGAAGAACAAACATGGAAAAA 964
      ||| | | ||||| | || || | | |||
Db      920 TTTATGTGGCATAACAAGAAATGTCATGGAAAAACAAGGCAGAAA 964

```

RESULT 7

ACF63388

ID ACF63388 standard; DNA; 3779 BP.

XX

AC ACF63388;

XX

DT 09-OCT-2003 (first entry)

XX

DE Human IBAT gene SEQ ID NO:110.

XX

KW Human; pharmacological; hypotensive; antilipaemic; vasotropic; laxative;
KW dermatological; antidepressant; tranquilliser; antiinflammatory; eczema;
KW antiulcer; antimigraine; neuroprotective; antiparkinsonian; analgesic;
KW gynaecological; virucide; vulnerary; antiarthritic; antipsoriatic; cold;
KW antimicrobial; cytostatic; litholytic; pathological disorder; depression;
KW abnormal appetite; hypertension; hypercholesterolaemia; hyperlipidaemia;
KW erectile dysfunction; anxiety; stress; inflammatory bowel syndrome;
KW ulcerative colitis; Crohn's disease; renal stone; gall stone; migraine;
KW constipation; headache; seizure; multiple sclerosis; polymyositis;
KW fibromyalgia; Parkinson's disease; amyotrophic lateral sclerosis; trauma;
KW chronic pain; pre-menstrual syndrome; sinusitis; carpal tunnel syndrome;

KW chronic fatigue syndrome; rosacea; arthritis; psoriasis; prostatitis;
KW inflammation; heart burn; infection; colon cancer; malignant melanoma;
KW skin disorder; gene; ds.
XX
OS Homo sapiens.
XX
PN WO2003006478-A1.
XX
PD 23-JAN-2003.
XX
PF 10-JUL-2002; 2002WO-US021664.
XX
PR 10-JUL-2001; 2001US-0303820P.
XX
PA (OLIG-) OLIGOS ETC INC.
XX
PI Dale RMK, Arrow A, Thompson T;
XX
DR WPI; 2003-221709/21.
XX
PT Composition with a modified oligonucleotide useful for treating a patient
PT with a pathological disorder such as abnormal appetite, hypertension,
PT eczema, anxiety, stress, and cancer.
XX
PS Claim 6; Page 130-132; 173pp; English.
XX
CC The present invention describes a composition (I) suitable for
CC administration in a mammal, which comprises a modified oligonucleotide
CC (II) of 7-75 nucleotides containing 7 or more contiguous ribose groups
CC linked by achiral 5'-3' internucleoside phosphate linkages, where the
CC modified oligonucleotide is complementary to a region of a gene
CC associated with a pathological disorder. Also described: (1) a
CC nutritional supplement comprising (II); and (2) a cosmetic composition
CC comprising (II), where the modified oligonucleotide is complementary to a
CC region of a gene associated with a skin disorder. (I) and (II) can have
CC hypotensive, antilipaemic, vasotropic, dermatological, antidepressant,
CC tranquilliser, antiinflammatory, antiulcer, laxative, antimigraine,
CC neuroprotective, antiparkinsonian, analgesic, gynaecological, virucide,
CC vulnerary, antiarthritic, antipsoriatic, antimicrobial, cytostatic and
CC litholytic activities. (I) can be used for treating a patient with a
CC pathological disorder selected from abnormal appetite, hypertension,
CC hypercholesterolaemia, hyperlipidaemia, erectile dysfunction, eczema,
CC depression, anxiety, stress, inflammatory bowel syndrome, ulcerative
CC colitis, Crohn's disease, renal stones, gall stones, constipation, colds,
CC migraine headache, seizure, multiple sclerosis, polymyositis, sinusitis,
CC fibromyalgia, Parkinson's disease, amyotrophic lateral sclerosis (ALS),
CC chronic pain, pre-menstrual syndrome, trauma, carpal tunnel syndrome,
CC chronic fatigue syndrome, rosacea, arthritis, psoriasis, prostatitis,
CC inflammation, heart burn, infection, poison ivy, colon cancer, malignant
CC melanoma, and malignant nasal polyps. The nutritional supplement is
CC useful for supplementing the diet of an individual, and the cosmetic
CC composition is useful for improving the appearance of the skin in an
CC individual with a skin disorder. ACF63279 to ACF63410 represent
CC nucleotide sequence given in the exemplification of the present invention
XX
SQ Sequence 3779 BP; 1117 A; 737 C; 799 G; 1126 T; 0 U; 0 Other;

Query Match 26.3%; Score 297.8; DB 7; Length 3779;
 Best Local Similarity 58.5%; Pred. No. 2.3e-82;
 Matches 518; Conservative 0; Mismatches 367; Indels 0; Gaps 0;

Qy 80 ATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGT 139
 || | || | || | | || | | || | || | || | || |
 Db 678 ATAACATCCTAAGTGTGGTCCTAAGTACGGTGTGACCATCCTGTTGGCCTTGGTGATGT 737

Qy 140 TCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCGCACATCAGGAGACCCTGGG 199
 |||| |||| |||| |||| || || | |||| | | | ||||
 Db 738 TCTCCATGGGATGCAACGTGGAAATCAAGAAATTTCTAGGGCACATAAAGCGGCCGTGGG 797

Qy 200 GCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGG 259
 |||| | || | | |||| |||| || |||| | ||||
 Db 798 GCATTGTGTGGCTTCCTCTGTGAGTTTGGAAATCATGCCCTCACAGGATTCATCCTGT 857

Qy 260 CCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCC 319
 | | | || | | || | || | || | | | || | || | || | || |
 Db 858 CGGTGGCCTTTGACATCCTCCCGCTCCAGGCCGTAGTGGTGTCTATTATAGGATGCTGCC 917

Qy 320 CGGGGGGCACCATCTCTAACATTTTACCTTCTGGGTGATGGAGATATGGATCTCAGCA 379
 | || | || | || | || | || | || | || | || | || | || |
 Db 918 CTGGAGGAAGTGCCTCCAATATCTTGGCCTATTGGGTGATGGCGACATGGACCTGAGCG 977

Qy 380 TCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATC 439
 |||| |||| | || |||| | | |||| |||| || || || | |
 Db 978 TCAGCATGACCACATGCTCCACACTGCTGCCCTCGGAATGATGCCGCTGTGCCTCCTTA 1037

Qy 440 TCTACACCTGGTCTTGGAGTCTTCAGCAGAATCTCACCATTTCCTTATCAGAACATAGGAA 499
 |||| || | || | || | || | || | || | || | || | || |
 Db 1038 TCTATACCAAAATGTGGGTGCACTCTGGGAGCATCGTAATTCCTATGATAACATAGGTA 1097

Qy 500 TTACCCTTGTGTGCCTGACCATTCTGTGGCCTTTGGTGTCTATGTGAATTACAGATGGC 559
 | || | | || | |||| | || | | | || | || | || | || |
 Db 1098 CATCTCTGGTTGCTCTCGTTGTTCTGTTTCCATTGGAATGTTTGTTAATCACAAATGGC 1157

Qy 560 CAAAACAATCCAAAATCATTTCTCAAGATTGGGGCCGTTGTTGGTGGGGTCCTCCTTCTGG 619
 | || | | || | || || || || || | | | || | || | || |
 Db 1158 CCCAAAAGCAAAGATCATACTTAAAATTGGGTCCATCGCGGGCGCCATCCTCATTGTGC 1217

Qy 620 TGGTCGCAGTTGCTGGTGTGGTCTTGGCGAAAGGATCTTGAATTCAGACATCACCTTC 679
 | | || | | || | | || | || | || | || | || | || |
 Db 1218 TCATAGCTGTGGTTGGAGGAATATTGTACCAAAGCGCCTGGATCATTGCTCCCAAAGTGT 1277

Qy 680 TGACCATCAGTTTCATCTTTCCTTTGATTGGCCATGTCACGGGTTTTCTGCTGGCACTTT 739
 || || | || |||| || || | | || |||| ||||
 Db 1278 GGATTATAGGAACAATATTTCTGTGGCGGGTTACTCCCTGGGGTTTCTTCTGGCTAGAA 1337

Qy 740 TTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCTTAGAACTGGAGCTCAGAATA 799
 || | | || | |||| | || | | || | || | || | || |
 Db 1338 TTGCTGGTCTACCTGGTACAGGTGCCGAACGGTTGCTTTTGAAACGGGGATGCAGAACA 1397

Qy 800 TTCAGATGTGCATCACCATGCTCCAGTTATCTTTCACTGCTGAGCACTTGGTCCAGATGT 859
 ||| | || |||| | || | || |||| |||| | | || |
 Db 1398 CGCAGCTATGTTCCACCATCGTTCAGCTCTCCTTCACTCCTGAGGAGCTCAATGTCGTAT 1457

Qy 860 TGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGATGGATTTCTTATTGTTGCAG 919

```

      | | | | | | | | | | | | | | | | | | | | | |
Db    1458 TCACCTTCCCGCTCATCTACAGCATTTCAGCTCGCCTTGCCGCAATATTCTTAGGAT 1517

Qy    920 CATATCAGACGTACAAGAGGAGATTGAAGAACAAACATGGAAAAA 964
      | | | | | | | | | | | | | | | | | | | | | |
Db    1518 TTTATGTGGCATAACAAGAAATGTCATGGAAAAACAAGGCAGAAA 1562

```

RESULT 8

ABZ20750

ID ABZ20750 standard; DNA; 3779 BP.

XX

AC ABZ20750;

XX

DT 28-MAR-2003 (first entry)

XX

DE Human ileal sodium-dependent bile acid transporter gene fragment #1.

XX

KW Human; ileal sodium-dependent bile acid transporter gene; SLC10A2; SNP;

KW single nucleotide polymorphism; chromosome 13q33; cardiant;

KW antiarteriosclerotic; antilipemic; gene; ds.

XX

OS Homo sapiens.

XX

FH	Key	Location/Qualifiers
FT	variation	replace(582,G)
FT		/*tag= a
FT	variation	replace(664,C)
FT		/*tag= b
FT	variation	replace(727,T)
FT		/*tag= c
FT	variation	replace(792,T)
FT		/*tag= d
FT	variation	replace(890,A)
FT		/*tag= e
FT	variation	replace(1073,A)
FT		/*tag= f
FT	variation	replace(1103,T)
FT		/*tag= g
FT	variation	replace(1384,T)
FT		/*tag= h
FT	variation	replace(1466,T)
FT		/*tag= i
FT	variation	replace(1484,C)
FT		/*tag= j
FT	variation	replace(1545,A)
FT		/*tag= k
FT	variation	replace(1646,T)
FT		/*tag= l
FT	variation	replace(1683,C)
FT		/*tag= m
FT	variation	replace(1765,C)
FT		/*tag= n

```

XX 11-APR-2002; 2002WO-GB001681.
XX
PR 17-APR-2001; 2001GB-00009296.
PR 19-APR-2001; 2001US-0284530P.
XX
PA (ASTR ) ASTRAZENECA AB.
PA (ASTR ) ASTRAZENECA UK LTD.
XX
PI Morten JEN;
XX
DR WPI; 2003-046927/04.
XX
PT Diagnosing polymorphism in SLC10A2 in a human for assessing the
PT pharmacogenetics of a drug for treating cardiovascular and hyperlipidemic
PT conditions, by determining the status of the human by reference to
PT polymorphism in SLC10A2.
XX
PS Claim 4; Page 19-20; 21pp; English.
XX
CC The present invention relates to a method of diagnosing polymorphisms in
CC SLC10A2 (human ileal sodium-dependent bile acid transporter gene) in a
CC human, which involves determining the status of the human by reference to
CC polymorphisms in SLC10A2. The method is useful for assessing the
CC pharmacogenetics of a drug acting at SLC10A2. The SLC10A2 gene
CC polymorphism is useful as a genetic marker in a linkage study. SLC10A2
CC drugs are also useful for treating cardiovascular (e.g. atherosclerosis)
CC and hyperlipidemic conditions. The SLC10A2 gene is found at chromosome
CC 13q33. The present sequence is a fragment of the gene of the invention
CC containing polymorphisms
XX
SQ Sequence 3779 BP; 1117 A; 737 C; 799 G; 1126 T; 0 U; 0 Other;

Query Match          26.3%;  Score 297.8;  DB 7;  Length 3779;
Best Local Similarity 58.5%;  Pred. No. 2.3e-82;
Matches 518;  Conservative 0;  Mismatches 367;  Indels 0;  Gaps 0;

Qy      80 ATGGAACCTGGAGCTCGTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGT 139
      || | ||| | || | | ||| || | || ||| || | ||||
Db      678 ATAACATCCTAAGTGTGGTCCTAAGTACGGTGCTGACCATCCTGTTGGCCTTGGTGATGT 737

Qy      140 TCTCTTTGGGATGTTCGGTGGAGATCCGGAAGCTGTGGTCGCACATCAGGAGACCCTGGG 199
      |||| ||||| ||||| ||| ||| | ||||| | | | ||||
Db      738 TCTCCATGGGATGCAACGTGGAAATCAAGAAATTTCTAGGGCACATAAAGCGGCCGTGGG 797

Qy      200 GCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGG 259
      |||| | || | | |||| ||||| ||||| | ||| | ||||
Db      798 GCATTTGTGTTGGCTTCCTCTGTCAAGTTTGGGAATCATGCCCTCACAGGATTCATCCTGT 857

Qy      260 CCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCC 319
      | | |||| | || |||| || | | |||| | | |||||
Db      858 CGGTGGCCTTTGACATCCTCCCGCTCCAGGCCGTAGTGGTGCTCATTATAGGATGCTGCC 917

Qy      320 CGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGATGGAGATATGGATCTCAGCA 379
      | || || | || || || || || || |||| |||| | |||| | |||
Db      918 CTGGAGGAACCTGCCTCCAATATCTTGGCCTATTGGGTGCGATGGCGACATGGACCTGAGCG 977

```


QY 380 TCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATC 439
 ||| |||| || || |||| || || |||| |||| |||| || ||| | |
 Db 978 TCAGCATGACCACATGCTCCACACTGCTTGCCCTCGGAATGATGCCGCTGTGCCTCCTTA 1037

QY 440 TCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATTCCCTTATCAGAACATAGGAA 499
 |||| ||| ||| || || |||| ||| | |||| ||| |
 Db 1038 TCTATACCAAAATGTGGGTCGACTCTGGGAGCATCGTAATCCCTATGATAACATAGGTA 1097

QY 500 TTACCCTTGTGTGCCTGACCATTCCCTGTGGCCTTTGGTGTCTATGTGAATTACAGATGGC 559
 | || || || |||| || |||| | | || || || ||||
 Db 1098 CATCTCTGGTTGCTCTCGTTGTTCTGTTCCATTGGAATGTTTGTAAATCACAAATGGC 1157

QY 560 CAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTTGGTGGGGTCCTCCTTCTGG 619
 | || || || || |||| || |||| || || || || || || ||
 Db 1158 CCCAAAAGCAAAGATCATACTTAAATGGGTCCATCGCGGGCGCCATCCTCATTGTGC 1217

QY 620 TGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGATCTTGAATTCAGACATCACCCCTTC 679
 | || || || || || | || || || || || || || || || ||
 Db 1218 TCATAGCTGTGGTTGGAGGAATATTGTACCAAAGCGCCTGGATCATTGCTCCCAAACGT 1277

QY 680 TGACCATCAGTTTCATCTTTCCCTTTGATTGGCCATGTCACGGGTTTTCTGCTGGCACTTT 739
 || || | || |||| || || | | || |||| ||||
 Db 1278 GGATTATAGGAACAATATTTCCCTGTGGCGGGTACTCCCTGGGGTTTCTTCTGGCTAGAA 1337

QY 740 TTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCCCTAGAACTGGAGCTCAGAATA 799
 || | | || || |||| || || || || || || || || || ||
 Db 1338 TTGCTGGTCTACCCTGGTACAGGTGCCGAACGGTTGCTTTTGAAACGGGGATGCAGAACA 1397

QY 800 TTCAGATGTGCATCACCATGCTCCAGTTATCTTTCACTGCTGAGCACTTGGTCCAGATGT 859
 ||| | || |||| || || || || |||| |||| | | || ||
 Db 1398 CGCAGCTATGTTCCACCATCGTTCAGCTCTCCTTCACTCCTGAGGAGCTCAATGTCGTAT 1457

QY 860 TGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGATGGATTTCTTATTGTTGCAG 919
 | | |||| || || | | |||| || || || || || || || ||
 Db 1458 TCACCTTCCCGCTCATCTACAGCATTTTCCAGCTCGCCTTTGCCGCAATATTCTTAGGAT 1517

QY 920 CATATCAGACGTACAAGAGGAGATTGAAGAACAACATGGAAAAA 964
 ||| | | |||| || || || || || || || || || || ||
 Db 1518 TTTATGTGGCATAACAAGAAATGTCATGGAAAAACAAGGCAGAAA 1562

RESULT 9

ADB58285

ID ADB58285 standard; DNA; 4269 BP.

XX

AC ADB58285;

XX

DT 04-DEC-2003 (first entry)

XX

DE Toxicity-related gene, SEQ ID 3311.

XX

KW Toxic; toxin; gene expression profile; hepatotoxicity; liver;

KW drug screening; toxicity assay; ds.

XX

OS Unidentified.

XX

PN WO2003064624-A2.

XX
PD 07-AUG-2003.
XX
PF 31-JAN-2003; 2003WO-US003194.
XX
PR 31-JAN-2002; 2002US-00060087.
PR 15-MAR-2002; 2002US-0364045P.
PR 15-MAR-2002; 2002US-0364055P.
PR 30-DEC-2002; 2002US-0436643P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
XX
DR WPI; 2003-689530/65.
XX
PT Predicting a toxic effect of a compound, useful in identifying toxicity
PT markers in liver tissues or cells for drug screening and toxicity assays,
PT comprises preparing gene expression profile of tissue or cells exposed to
PT the compound.
XX
PS Claim 1; SEQ ID NO 3311; 1156pp; English.
XX
CC The present invention relates to a method for predicting a toxic effect
CC of a compound. The method comprises preparing a gene expression profile
CC of a tissue or cell sample exposed to the compound, and comparing the
CC gene expression profile to a database comprising SEQ ID 1-4925, where
CC differential expression of the gene indicates at least one toxic effect.
CC The method is useful for predicting at least one toxic effect of a
CC compound, predicting hepatotoxicity or the progression of a toxic effect
CC of a compound, identifying an agent that modulates the onset or
CC progression of a toxic response, predicting the cellular pathways that a
CC compound modulates in a cell, and identifying an agent that modulates at
CC least one activity of a protein. The method and compositions of the
CC present invention using a database of genes having liver toxin-induced
CC differential expression, are useful in identifying toxicity markers in
CC liver tissues or cells for drug screening and toxicity assays. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4269 BP; 1315 A; 780 C; 850 G; 1324 T; 0 U; 0 Other;

Query Match 25.9%; Score 293.2; DB 9; Length 4269;
Best Local Similarity 59.1%; Pred. No. 7.1e-81;
Matches 502; Conservative 0; Mismatches 348; Indels 0; Gaps 0;

```

Qy      109 GTGTCCACTGTGATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGG 168
      ||| ||| | | | | || | |||| ||| ||| || | |||| |||
Db      216 GTGCTCACCATTCTTCTAGCCATGGTGATGTTTTCTATGGGGTGCAATGTGGAATCAAC 275

Qy      169 AAGCTGTGGTCGCACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTT 228
      ||| | | |||| | | | || |||| || | |||| |||||
Db      276 AAGTTCCTAGGACACATAAAGCGGCCATGGGGCATCTTCGTGGGCTTCCTCTGTCAGTTT 335

Qy      229 GGGCTCATGCCTTTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAA 288
      || | |||| || | |||| | | |||| | | || | | || |||

```

Db 336 GGAATCATGCCTCTCACAGGATTTATCCTGTCTGTGGCCTCTGGCATCCTTCCTGTGCAG 395
 Qy 289 GCTATTGCTGTTCTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACC 348
 ||| | | | | | | | | | | | | | | | | | | | | |
 Db 396 GCTGTGGTGGTGCTAATTATGGGTTGCTGCCCTGGAGGAACCTGGCTCCAATATCCTGGCC 455
 Qy 349 TTCTGGGTTGATGGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCC 408
 | | | | | | | | | | | | | | | | | | | | | |
 Db 456 TATTGGATAGATGGTGACATGGACCTCAGTGTTAGCATGACCACTTGCTCCACACTGCTT 515
 Qy 409 GCCCTGGGAATGATGCCACTCTGCATTTATCTCTACACCTGGTCCTGGAGTCTTCAGCAG 468
 || | | | | | | | | | | | | | | | | | | | |
 Db 516 GCTCTTGAATGATGCCCTTTGCCTCTTCATCTATAACCAAGATGTGGGTTGACTCAGGA 575
 Qy 469 AATCTCACCATTCCCTTATCAGAACATAGGAATTACCCTTGTGTGCCTGACCATTCTCTGTG 528
 | | | | | | | | | | | | | | | | | | | | | |
 Db 576 ACGATTGTGATCCCCACGATAGCATTGGCATTCTCTGTTGCGCTTGTTATTCCTGTT 635
 Qy 529 GCCTTTGGTGTCTATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATT 588
 || | | | | | | | | | | | | | | | | | | | |
 Db 636 TCCATTGGAATGTTGTAAATCACAAATGGCCCCAAAAGCGAAGATTATACTTAAAAAT 695
 Qy 589 GGGGCCCGTTGTTGGTGGGGTCCTCCTTCTGGTGGTCGCAGTTGCTGGTGTGGTCCTGGCG 648
 || | | | | | | | | | | | | | | | | | | | |
 Db 696 GGATCCATCGCAGGTGCAATTCTCATTGTGCTCATAGCTGTGGTTGGAGGAATACTGTAC 755
 Qy 649 AAAGGATCTTGAATTTCAGACATCACCCCTTCTGACCATCAGTTTCATCTTTCCCTTTGATT 708
 || | | | | | | | | | | | | | | | | | | | |
 Db 756 CAAAGTGCCCTGGATCATTGAACCCAAACTATGGATTATAGGAACAATATTTCCATAGCT 815
 Qy 709 GGCCATGTCACGGGTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGG 768
 ||| | | | | | | | | | | | | | | | | | | |
 Db 816 GGCTACAGCCTTGGTTTCTTCCTGGCTAGACTAGCTGGTCAACCCTGGTACAGGTGCCGA 875
 Qy 769 ACAATTTCCCTAGAAACTGGAGCTCAGAATATTAGATGTGCATCACCATGCTCCAGTTA 828
 ||| | | | | | | | | | | | | | | | | | | |
 Db 876 ACAGTTGCCTTGGAAACTGGAATGCAGAACTCAACTGTGTTCCACCATTGTACAACCTC 935
 Qy 829 TCTTTCACTGCTGAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTC 888
 || | | | | | | | | | | | | | | | | | | | |
 Db 936 TCCTTTAGCCCTGAGGATCTCAACCTTGTGTTCACCTTCCCACTCATCTATACTGTTTTC 995
 Qy 889 CAGCTGATAGATGGATTTCTTATTGTTGCAGCATATCAGACGTACAAGAGGAGATTGAAG 948
 |||| | | | | | | | | | | | | | | | | | |
 Db 996 CAGCTCGTCTTTGCAGCAATAATATTAGGAATGTATGTCACATACAAGAAATGTCATGGA 1055
 Qy 949 AACAAACATG 958
 || | | | |
 Db 1056 AAAAATGATG 1065

RESULT 10

ADB52825

ID ADB52825 standard; DNA; 4269 BP.

XX

AC ADB52825;

XX

DT 04-DEC-2003 (first entry)
XX
DE Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:3367.
XX
KW toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;
KW toxicity marker; toxicity progression; drug screening;
KW primary rat hepatocyte toxicity modelling; gene; ds.
XX
OS Rattus norvegicus.
XX
PN WO2003065993-A2.
XX
PD 14-AUG-2003.
XX
PF 04-FEB-2003; 2003WO-US003482.
XX
PR 04-FEB-2002; 2002US-0353171P.
PR 13-MAR-2002; 2002US-0363534P.
PR 08-APR-2002; 2002US-0370248P.
PR 10-APR-2002; 2002US-0371134P.
PR 10-APR-2002; 2002US-0371135P.
PR 10-APR-2002; 2002US-0371150P.
PR 11-APR-2002; 2002US-0371413P.
PR 19-APR-2002; 2002US-0373601P.
PR 19-APR-2002; 2002US-0373602P.
PR 22-APR-2002; 2002US-0374139P.
PR 08-MAY-2002; 2002US-0378370P.
PR 09-MAY-2002; 2002US-0378652P.
PR 09-MAY-2002; 2002US-0378653P.
PR 09-MAY-2002; 2002US-0378665P.
PR 09-JUL-2002; 2002US-0394230P.
PR 09-JUL-2002; 2002US-0394253P.
PR 04-SEP-2002; 2002US-0407688P.
PR 28-JAN-2003; 2003US-0442900P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;
PI Elashoff M;
XX
DR WPI; 2003-731472/69.
XX
PT Determining if a compound induces a toxic effect on a tissue or cell, for
PT identifying hepatotoxic compounds, comprises comparing a gene expression
PT profile of a tissue or cell sample to a database of Tox mean and non-Tox
PT mean values.
XX
PS Claim 44; SEQ ID NO 3367; 874pp; English.
XX
CC The present invention describes a method for determining whether a
CC compound induces a toxic effect on a tissue or cell. The method comprises
CC preparing a gene expression profile of a tissue or cell sample exposed to
CC the compound, and comparing the gene expression profile to a database
CC comprising data or information on the Tox mean and non-Tox mean value.
CC The method is useful for predicting or identifying at least one toxic
CC effect, particularly hepatotoxicity, of a test or unknown compound. The
CC genes listed in the specification are useful as diagnostic or toxicity

CC markers for the prediction or identification of the physiological state
CC of tissue or cell sample that has been exposed to a compound, or to
CC identify or predict the toxic effects of a compound or an agent. These
CC may also be used as markers for monitoring toxicity progression or for
CC drug screening. The present sequence represents a primary rat hepatocyte
CC toxicity modelling related gene sequence from the present invention.

XX

SQ Sequence 4269 BP; 1315 A; 780 C; 850 G; 1324 T; 0 U; 0 Other;

Query Match 25.9%; Score 293.2; DB 9; Length 4269;
Best Local Similarity 59.1%; Pred. No. 7.1e-81;
Matches 502; Conservative 0; Mismatches 348; Indels 0; Gaps 0;

```
Qy      109 GTGTCCACTGTGATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGG 168
      ||| ||| | | | | | | | | | | | | | | | | | | | | |
Db      216 GTGCTCACCATTCTTCTAGCCATGGTGTATGTTTTCTATGGGGTGCAATGTGGAAATCAAC 275

Qy      169 AAGCTGTGGTCGCACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTT 228
      ||| | | | | | | | | | | | | | | | | | | | | |
Db      276 AAGTTCCTAGGACACATAAAGCGGCCATGGGGCATCTTCGTGGGCTTCTCTGTCAAGTTT 335

Qy      229 GGGCTCATGCCTTTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAA 288
      || | | | | | | | | | | | | | | | | | | | | |
Db      336 GGAATCATGCCTCTCACAGGATTTATCCTGTCTGTGGCCTCTGGCATCCTTCTGTGCAG 395

Qy      289 GCTATTGCTGTTCTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTTAACATTTTCACC 348
      ||| | | | | | | | | | | | | | | | | | | | | |
Db      396 GCTGTGGTGGTGCTAATTATGGGTTGCTGCCCTGGAGGAACGGCTCCAATATCCTGGCC 455

Qy      349 TTCTGGGTTGATGGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCC 408
      | | | | | | | | | | | | | | | | | | | | | | |
Db      456 TATTGGATAGATGGTGACATGGACCTCAGTGTTAGCATGACCACTTGCTCCACACTGCTT 515

Qy      409 GCCCTGGGAATGATGCCACTCTGCATTTATCTCTACACCTGGTCCTGGAGTCTTCAGCAG 468
      || | | | | | | | | | | | | | | | | | | | | |
Db      516 GCTCTTGAATGATGCCCCCTTTCCTCTTATCTATACCAAGATGTGGGTTGACTCAGGA 575

Qy      469 AATCTCACCATTCCCTTATCAGAACATAGGAATTACCCTTGTGTGCCTGACCATTCCCTGTG 528
      | | | | | | | | | | | | | | | | | | | | | |
Db      576 ACGATTGTGATCCCCTACGATAGCATTGGCATTCTCTGTTGCGCTTGTATTCCTGTT 635

Qy      529 GCCTTTGGTGTCTATGTGAATTACAGATGGCCAAAACAATCCAAAATCATCTCAAGATT 588
      || |||| | | ||| ||| ||| |||| | | | | | | | |
Db      636 TCCATTGGAATGTTTGTAATCACAAATGGCCCCAAAAGCGAAGATTATACTTAAAATT 695

Qy      589 GGGGCCGTTGTTGGTGGGGTCCTCCTTCTGGTGGTGCAGTTGCTGGTGTGGTTCCTGGCG 648
      || || | | | | | | | | | | | | | | | | | | |
Db      696 GGATCCATCGCAGGTGCAATTCTCATTGTGCTCATAGCTGTGGTTGGAGGAATACTGTAC 755

Qy      649 AAAGGATCTTGAATTCAGACATCACCCCTTCTGACCATCAGTTTCATCTTTCCTTTGATT 708
      || | | | | | | | | | | | | | | | | | | | | |
Db      756 CAAAGTGCCTGGATCATTGAACCCAACTATGGATTATAGGAACAATATTTCTATAGCT 815

Qy      709 GGCCATGTCACGGGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGG 768
      ||| | | | | | | | | | | | | | | | | | | | |
Db      816 GGCTACAGCCTTGGTTTCTTCTGCTAGCTAGCTGGTCAACCCTGGTACAGGTGCCGA 875
```

Qy 769 ACAATTTTCCTTAGAAACTGGAGCTCAGAATATTCAGATGTGCATCACCATGCTCCAGTTA 828
 ||| || |||| ||||| |||| | ||| |||| ||||| | || |
 Db 876 ACAGTTGCCTTGGAACCTGGAATGCAGAACACTCAACTGTGTCCACCATTGTACAACCTC 935

Qy 829 TCTTTCAGTGTGAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTC 888
 || || | |||| | | || |||| | ||||| |||| | |||
 Db 936 TCCTTTAGCCCTGAGGATCTCAACCTTGTGTTACCTTCCCACTCATCTATACTGTTTTTC 995

Qy 889 CAGCTGATAGATGGATTTCTTATTGTTGCAGCATATCAGACGTACAAGAGGAGATTGAAG 948
 |||| | || | || | || | || ||||| |
 Db 996 CAGCTCGTCTTTGCAGCAATAATATTAGGAATGTATGTCACATACAAGAAATGTCATGGA 1055

Qy 949 AACAAACATG 958
 || || |||
 Db 1056 AAAAATGATG 1065

RESULT 11

ABK63719

ID ABK63719 standard; cDNA; 1663 BP.

XX

AC ABK63719;

XX

DT 18-JUN-2002 (first entry)

XX

DE Rat sequence differentially expressed in response to a hepatotoxin #1626.

XX

KW Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;

KW differential expression; centrilobular necrosis; steatosis.

XX

OS Rattus norvegicus.

XX

PN WO200210453-A2.

XX

PD 07-FEB-2002.

XX

PF 30-JUL-2001; 2001WO-US023872.

XX

PR 31-JUL-2000; 2000US-0222040P.

PR 02-NOV-2000; 2000US-0244880P.

PR 11-MAY-2001; 2001US-0290029P.

PR 15-MAY-2001; 2001US-0290645P.

PR 22-MAY-2001; 2001US-0292336P.

PR 06-JUN-2001; 2001US-0295798P.

PR 13-JUN-2001; 2001US-0297457P.

PR 19-JUN-2001; 2001US-0298884P.

PR 09-JUL-2001; 2001US-0303459P.

XX

PA (GENE-) GENE LOGIC INC.

XX

PI Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;

XX

DR WPI; 2002-241625/29.

XX

PT Predicting toxic effects of compounds or the progression of these toxic
 PT effects by determining the changes in gene expression in tissues or cells
 PT exposed to the toxin and comparing these to gene expression in unexposed

PT tissues or cells.

XX

PS Claim 1; SEQ ID NO 1626; 239pp; English.

XX

CC The invention relates to methods for predicting toxic effects of
CC compounds or the progression of these toxic effects by determining the
CC global changes in gene expression in tissues or cells exposed to the
CC toxin and comparing these to gene expression in unexposed tissues or
CC cells. Also included are methods of predicting at least one toxic effect
CC of a compound or progression of a toxic effect, preferably the
CC hepatotoxicity of a compound, comprising detecting the level of
CC expression in a tissue or cell sample exposed to the compound of two or
CC more genes listed in the specification, where differential expression of
CC the genes is indicative of at least one toxic effect or progression. The
CC method can also be used to identify an agent which modulates the toxic
CC response and predict cellular pathways that a compound modulates in a
CC cell. The methods utilise a set of at least two probes (on a solid
CC support in kit form), where each of the probes comprises a sequence that
CC specifically hybridises to a gene listed in the specification, a computer
CC system comprising a database containing information identifying the
CC expression level in a tissue or cell sample exposed to a hepatotoxin of a
CC set of genes comprising at least two genes listed in the specification,
CC and a user interface to view the information used to present information
CC identifying the expression level in a tissue or cell of at least one gene
CC listed in the specification. The method is useful for elucidating global
CC changes in gene expression and for identifying toxicity markers in
CC tissues or cell exposed to a known toxin. The genes may be used as
CC toxicity markers in drug screening and toxicity assays. The genes and
CC gene expression information may be used as diagnostic markers for the
CC prediction or identification of the physiological state of tissue or cell
CC sample that has been exposed to a compound or agent. Hepatotoxicity is
CC characterised by centrilobular necrosis and steatosis. The present
CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene
CC which is differentially expressed in response to a hepatotoxic agent

XX

SQ Sequence 1663 BP; 450 A; 460 C; 325 G; 428 T; 0 U; 0 Other;

Query Match 16.2%; Score 183.2; DB 6; Length 1663;

Best Local Similarity 53.6%; Pred. No. 1.9e-46;

Matches 430; Conservative 0; Mismatches 363; Indels 9; Gaps 2;

Qy 119 TGATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGT 178
| ||| || ||| |||| |||| |||| || || |||| || || ||| |

Db 219 TAATGTTGCTGCTTATCATGCTCTCACTGGGCTGCACCATGGAATTCAGCAAGATCAAGG 278

Qy 179 CGCACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGGCTCATGC 238
| ||| | ||| ||| || | || | || | ||||| |||||

Db 279 CTCACTTGTGGAAGCCCAAAGGGGTGATCGTTGCCTTGGTGGCCAGTTTGGCATCATGC 338

Qy 239 CTTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTG 298
| | | |||| || || | || | |||| |||| | |||| | ||

Db 339 CCCTCGCTGCTTTTCTTCTCGGCAAGATCTTTCACCTGAGCAACATTGAAGCTCTGGCCA 398

Qy 299 TTCTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTG 358
| ||||| ||||| || |||| | | | || | ||||| |

Db 399 TCCTCATCTGTGGCTGCTCTCCCGGGGGGAACCTTGTCACCTCTTCACCTGGCCATGA 458

Qy 359 ATGGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAA 418
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 459 AGGGGGACATGAACCTCAGCATCGTGATGACCACCTGCTCCAGCTTCAGTGCCTTGGGCA 518
 419 TGATGCCACTCTGCATTTATCTCTACACC---TGGTCCTGGAGTCTTCAGCAGAATCTCA 475
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 519 TGATGCCACTCCTCTTATACGTCTACAGCAAAGGCATCTACGATGGAGACCTTAAGGACA 578
 476 CCATTCCTTATCAGAACATAGGAATTACCCTTGTGTGCCTGACCATTCTGTGGCCTTTG 535
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 579 AGGTGCCCTACAAAGGCATTATGATATCACTAGTCATAGTTCTCATTCTGCACCATAG 638
 536 GTGTCTATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATTGGGGCCG 595
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 639 GGATCGTCCTCAAGTCCAAAAGGCCACACTATGTACCCTACATCCTCAAGGGAGGCATGA 698
 596 TTGTTGGTGGGGTCCTCCTTCTGGTGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGAT 655
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 699 TCATCACCTTCCTCCTCTCTGTGGCTGTCACAGCCCTCTCTGTCTCATCAATGTGGGCAACA 758
 656 CTTGGAATTCAGACATCAC-----CCTTCTGACCATCAGTTTCATCTTTCCTTTGATTG 709
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 759 GCATCATGTTTCGTTCATGACACCACACTTACTGGCTACCTCCTCCCTGATGCCCTTCTCTG 818
 710 GCCATGTCACGGGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGGA 769
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 819 GCTTTCTGATGGGTTACATTCTCTGCTCTCTTCCAACCTCAATCCAAGCTGCAGACGCA 878
 770 CAATTTCTTAGAACTGGAGCTCAGAATATTCAGATGTGCATCACCATGCTCCAGTTAT 829
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 879 CCATCAGCATGGAAACAGGATTCCAAAACATTCAACTCTGTTCTACCATCCTCAATGTGA 938
 830 CTTTCACTGCTGAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTCC 889
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 939 CCTTCCCCCTGAAGTCATTGGGCCACTTTTCTTCTTCTCCTCTCTACATGATTTTCC 998
 890 AGCTGATAGATGGATTTCTTAT 911
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 999 AGCTTGAGAAGGACTTCTCAT 1020

RESULT 12

ADB58234

ID ADB58234 standard; DNA; 1663 BP.

XX

AC ADB58234;

XX

DT 04-DEC-2003 (first entry)

XX

DE Toxicity-related gene, SEQ ID 3260.

XX

KW Toxic; toxin; gene expression profile; hepatotoxicity; liver;

KW drug screening; toxicity assay; ds.

XX

OS Unidentified.

XX

PN WO2003064624-A2.

KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumour; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS Homo sapiens.
XX
PN WO200229103-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US030589.
XX
PR 02-OCT-2000; 2000US-0237054P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX
DR WPI; 2002-426119/45.
XX
PT Diagnosing and detecting the progression of liver cancer, hepatocellular
PT carcinoma or metastatic liver tumor in a patient, involves detecting the
PT level of expression of two or more genes in a liver tissue sample.
XX
PS Claim 1; SEQ ID NO 2176; 298pp; English.
XX
CC The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumour in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cytostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1580 BP; 400 A; 434 C; 341 G; 405 T; 0 U; 0 Other;

Query Match 15.3%; Score 173.6; DB 6; Length 1580;
Best Local Similarity 51.9%; Pred. No. 2e-43;
Matches 445; Conservative 0; Mismatches 404; Indels 9; Gaps 2;

Qy 119 TGATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGT 178
| | | | | | | | | | | | | | | | | | | | | |
Db 180 TCATGTTGTTCTTCATCATGCTCTCGCTGGGCTGCACCATGGAGTTCAGCAAGATCAAGG 239
Qy 179 CGCACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGC 238
| | | | | | | | | | | | | | | | | | | | | |
Db 240 CTCACCTATGGAAGCCTAAAGGGCTGGCCATCGCCCTGGTGGCACAGTATGGCATCATGC 299
Qy 239 CTTTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTG 298

Db	300	CCCTCACGGCCTTTGTGCTGGGCAAGGTCTTCCGGCTGAAGAACATTGAGGCACTGGCCA	359
Qy	299	TTCTCATCATGGGCTGCTGCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTG	358
Db	360	TCTTGGTCTGTGGCTGCTCACCTGGAGGGAACCTGTCCAATGTCTTCAGTCTGGCCATGA	419
Qy	359	ATGGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAA	418
Db	420	AGGGGGACATGAACCTCAGCATTTGTGATGACCACCTGCTCCACCTTCTGTGCCCTTGGCA	479
Qy	419	TGATGCCACTCTGCATTTATCTCTACACC---TGGTCCTGGAGTCTTCAGCAGAATCTCA	475
Db	480	TGATGCCTCTCCTCCTGTACATCTACTCCAGGGGGATCTATGATGGGGACCTGAAGGACA	539
Qy	476	CCATTTCCTTATCAGAACATAGGAATTACCCTTGTGTGCCTGACCATTCTGTGGCCTTTG	535
Db	540	AGGTGCCCTATAAAGGCATCGTGATATCACTGGTCCTGGTTCTCATTCTTGCACCATAG	599
Qy	536	GTGTCTATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATTGGGGCCG	595
Db	600	GGATCGTCCTCAAATCCAAACGGCCACAATACATGCGCTATGTCATCAAGGGAGGGATGA	659
Qy	596	TTGTTGGTGGGGTCCTCCTTCTGGTGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGAT	655
Db	660	TCATCATTCTCTTGTGCAGTGTGGCCGTCACAGTTCTCTCTGCCATCAATGTGGGGAAGA	719
Qy	656	CTTGAATTTCAGACATCAC-----CCTTCTGACCATCAGTTTCATCTTTCTTTGATTG	709
Db	720	GCATCATGTTTGCCATGACACCACTCTTGATTGCCACCTCCTCCCTGATGCCTTTTATTG	779
Qy	710	GCCATGTCACGGGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGGA	769
Db	780	GCTTTCTGCTGGGTTATGTTCTCTCTGCTCTCTTCTGCCTCAATGGACGGTGCAGACGCA	839
Qy	770	CAATTTCTTTAGAACTGGAGCTCAGAATATTCAGATGTGCATCACCATGCTCCAGTTAT	829
Db	840	CTGTCAGCATGGAGACTGGATGCCAAAATGTCCAACCTCTGTTCCACCATCCTCAATGTGG	899
Qy	830	CTTTCACTGCTGAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTCC	889
Db	900	CCTTTCCACCTGAAGTCATTGGACCACTTTTCTTCTTTCCCCTCCTCTACATGATTTTCC	959
Qy	890	AGCTGATAGATGGATTTCTTATTGTTGCAGCATATCAGACGTACAAGAGGAGATTGAAGA	949
Db	960	AGCTTGGAGAAGGGCTTCTCCTCATTGCCATATTTTGGTGCTATGAGAAATTCAAGACTC	1019
Qy	950	ACAAACATGGAAAAAGA	967
Db	1020	CCAAGGATAAAACAAAAA	1037

RESULT 14

AAD56518

ID AAD56518 standard; DNA; 1580 BP.

XX

AC AAD56518;

XX
 DT 27-AUG-2003 (first entry)
 XX
 DE Human sodium/bile acid cotransporter, 8587 DNA.
 XX
 KW Human; cardiovascular disorder; coronary artery disease; bradycardia;
 KW restenosis; cardiac hypertrophy; ischaemia reperfusion injury; angina;
 KW arteriosclerosis; coronary artery ligation; rheumatic heart disease;
 KW heart failure; hypertension; cardiomyopathy; myocardial infarction;
 KW arterial inflammation; microembolism; atherosclerosis; endocarditis;
 KW vascular heart disease; valvular disease; arrhythmia; gene therapy;
 KW sinus node dysfunction; sodium-bile acid cotransporter; gene; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 83. .1132
 FT /*tag= a
 FT /product= "Human sodium/bile acid cotransporter protein"
 XX
 PN WO2003039341-A2.
 XX
 PD 15-MAY-2003.
 XX
 PF 05-NOV-2002; 2002WO-US035538.
 XX
 PR 05-NOV-2001; 2001US-0339582P.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Logan TJ, Chun M, Galvin KM;
 XX
 DR WPI; 2003-441437/41.
 DR P-PSDB; AAE37351.
 XX
 PT Treating a subject having a cardiovascular disorder, e.g. angina,
 PT arrhythmia, or restenosis, comprises administering a 139, 258, 1261,
 PT 1486, 2398, 2414, 7660, 8587, 10183, 10550, 12680, 17921, 32248, 60489 or
 PT 93804 modulator.
 XX
 PS Disclosure; Page 109; 124pp; English.
 XX
 CC The invention relates to methods and compositions for treating a subject
 CC having a cardiovascular disorder using 139, 258, 1261, 1486, 2398, 2414,
 CC 7660, 8587, 10183, 10550, 12680, 17921, 32248, 60489 or 93804 modulator.
 CC The invention is useful for treating a cardiovascular disorder, including
 CC arteriosclerosis, atherosclerosis, vascular wall remodeling, restenosis,
 CC cardiac hypertrophy, ischaemia reperfusion injury, arterial inflammation,
 CC ventricular remodelling, rapid ventricular pacing, tachycardia, coronary
 CC microembolism, bradycardia, pressure overload, aortic bending, coronary
 CC artery ligation, vascular heart disease, valvular disease, including but
 CC not limited to, valvular degeneration caused by calcification, rheumatic
 CC heart disease, endocarditis, or complications of artificial valves;
 CC atrial fibrillation, long-QT syndrome, congestive heart failure, sinus
 CC node dysfunction, angina, heart failure, hypertension, atrial flutter,
 CC atrial fibrillation, pericardial disease, including but not limited to
 CC pericardial effusion and pericarditis, cardiomyopathies (e.g. dilated

CC cardiomyopathy or idiopathic cardiomyopathy), myocardial infarction,
CC coronary artery disease, coronary artery spasm, ischaemic disease,
CC arrhythmia, sudden cardiac death, and cardiovascular developmental
CC disorders. The invention is also useful in gene therapy. The present
CC sequence is human sodium/bile acid cotransporter DNA. This sequence is
CC used to illustrate the method of the invention

XX

SQ Sequence 1580 BP; 400 A; 434 C; 341 G; 405 T; 0 U; 0 Other;

Query Match 15.3%; Score 173.6; DB 7; Length 1580;
Best Local Similarity 51.9%; Pred. No. 2e-43;
Matches 445; Conservative 0; Mismatches 404; Indels 9; Gaps 2;

```
Qy      119 TGATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGT 178
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      180 TCATGTTGTTCTTCATCATGCTCTCGCTGGGCTGCACCATGGAGTTCAGCAAGATCAAGG 239

Qy      179 CGCACATCAGGAGACCCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGC 238
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      240 CTCACTTATGGAAGCCTAAAGGGCTGGCCATCGCCCTGGTGGCACAGTATGGCATCATGC 299

Qy      239 CTTTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTG 298
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      300 CCCTCACGGCCTTTGTGCTGGGCAAGGTCTTCCGGCTGAAGAACATTGAGGCACTGGCCA 359

Qy      299 TTCTCATCATGGGCTGCTGCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTG 358
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      360 TCTTGGTCTGTGGCTGCTCACCTGGAGGGAACCTGTCCAATGTCTTCAGTCTGGCCATGA 419

Qy      359 ATGGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAA 418
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      420 AGGGGGACATGAACCTCAGCATTTGTGATGACCACCTGCTCCACCTTCTGTGCCCTTGGCA 479

Qy      419 TGATGCCACTCTGCATTTATCTCTACACC---TGGTCCTGGAGTCTTCAGCAGAATCTCA 475
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      480 TGATGCCTCTCCTCCTGTACATCTACTCCAGGGGGATCTATGATGGGGACCTGAAGGACA 539

Qy      476 CCATTCCCTTATCAGAACATAGGAATTACCCTTGTGTGCCTGACCATTCTGTGGCCTTTG 535
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      540 AGGTGCCCTATAAAGGCATCGTGATATCACTGGTCCTGGTTCTCATTCTTGCACCATAG 599

Qy      536 GTGTCTATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATTGGGGCCG 595
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      600 GGATCGTCCTCAAATCCAAACGCCACAATACATGCGCTATGTCATCAAGGGAGGGATGA 659

Qy      596 TTGTTGGTGGGGTCCTCCTTCTGGTGGTFCGAGTTGCTGGTGTGGTCCTGGCGAAAGGAT 655
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      660 TCATCATTCTCTTGTGCAGTGTGGCCGTACAGTTCTCTCTGCCATCAATGTGGGGAAGA 719

Qy      656 CTTGGAATTCAGACATCAC-----CCTTCTGACCATCAGTTTCATCTTTCCTTTGATTG 709
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      720 GCATCATGTTTGCCATGACACCACTCTTGATTGCCACCTCCTCCCTGATGCCTTTTATTG 779

Qy      710 GCCATGTCACGGGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGGA 769
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      780 GCTTTCTGCTGGGTTATGTTCTCTGCTCTCTCTGCTCAATGGACGGTGCAGACGCA 839
```

Qy 770 CAATTCCTTAGAACTGGAGCTCAGAATATTCAGATGTGCATCACCATGCTCCAGTTAT 829
 | | | | | | | | | | | | | | | | | | | | | |
 Db 840 CTGTCAGCATGGAGACTGGATGCCAAAATGTCCAACCTCTGTTCCACCATCCTCAATGTGG 899
 Qy 830 CTTTCACTGCTGAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTCC 889
 | | | | | | | | | | | | | | | | | | | | | |
 Db 900 CCTTTCACCTGAAGTCATTGGACCACTTTTCTTCTTTCCCCTCCTCTACATGATTTTCC 959
 Qy 890 AGCTGATAGATGGATTCTTATTGTTGCAGCATATCAGACGTACAAGAGGAGATTGAAGA 949
 | | | | | | | | | | | | | | | | | | | | | |
 Db 960 AGCTTGGAGAAGGGCTTCTCCTCATTGCCATATTTTGGTGCTATGAGAAATTCAAGACTC 1019
 Qy 950 ACAAACATGGAAAAAAGA 967
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1020 CCAAGGATAAAACAAAAA 1037

RESULT 15

ACC51213

ID ACC51213 standard; cDNA; 1580 BP.

XX

AC ACC51213;

XX

DT 16-JUN-2003 (first entry)

XX

DE Human Plk-1 related cDNA sequence hmft-1603 SEQ ID NO:98.

XX

KW Human; hepatoblastoma; cancer detection probe; cancer; detection;
 KW hepatocellular carcinoma; hereditary non-polyposis colorectal cancer;
 KW desmoid tumour; anaplastic thyroid carcinoma; Wilm's tumour; tumour;
 KW Plk-1; polo-like kinase-1; gene; ss.

XX

OS Homo sapiens.

XX

PN WO2003018807-A1.

XX

PD 06-MAR-2003.

XX

PF 26-AUG-2002; 2002WO-JP008580.

XX

PR 24-AUG-2001; 2001JP-00255225.

XX

PA (HISM) HISAMITSU PHARM CO LTD.

PA (CHIB-) CHIBA PREFECTURE.

XX

PI Nakagawara A;

XX

DR WPI; 2003-268424/26.

XX

PT Nucleic acid sequences differently expressed between hepatoblastoma and
 PT normal liver tissue, are useful for cancer detection and diagnosis.

XX

PS Claim 4; Page 156-157; 180pp; Japanese.

XX

CC The present invention describes nucleic acid sequences (I) having a
 CC different degree of expression in hepatoblastoma from their expression in
 CC normal liver tissue. ACC51116 to ACC51219 represents specifically claimed

CC examples of (I). Also described: (1) nucleic acids stringently
 CC hybridising to (I); (2) cancer detection probes containing one or more of
 CC 104 listed sequences (II, see ACC51116 to ACC51219), including the 79 (I,
 CC see ACC51116 to ACC51194), or partial sequences derived from them; (3)
 CC PCR primers for cancer detection based on sequences (II); (4) marker
 CC proteins for cancer detection, encoded by (II); (5) diagnostic reagents
 CC for cancer diagnosis, containing (II) or their partial sequences. The
 CC nucleic acid sequences are useful in the detection and diagnosis of
 CC cancers including liver, colon, breast, kidney, bladder, ovary and
 CC thyroid cancer, especially for hepatoblastoma, hepatocellular carcinoma,
 CC hereditary non-polyposis colorectal cancer, desmoid tumour, anaplastic
 CC thyroid carcinoma and Wilm's tumour. They are also used as markers for
 CC predicting the prognosis of these tumours. ACC51220 to ACC51233 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC nucleic acid sequences given in ACC51116 to ACC51219 are related to human
 CC Plk-1 (polo-like kinase-1), which is located on chromosome 16p12

XX

SQ Sequence 1580 BP; 400 A; 434 C; 341 G; 405 T; 0 U; 0 Other;

Query Match 15.3%; Score 173.6; DB 7; Length 1580;
 Best Local Similarity 51.9%; Pred. No. 2e-43;
 Matches 445; Conservative 0; Mismatches 404; Indels 9; Gaps 2;

Qy	119	TGATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGT	178
Db	180	TCATGTTGTTCTTCATCATGCTCTCGCTGGGCTGCACCATGGAGTTCAGCAAGATCAAGG	239
Qy	179	CGCACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGC	238
Db	240	CTCACTTATGGAAGCCTAAAGGGCTGGCCATCGCCCTGGTGGCACAGTATGGCATCATGC	299
Qy	239	CTTTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTG	298
Db	300	CCCTCACGGCCTTTGTGCTGGGCAAGGTCTTCCGGCTGAAGAACATTGAGGCACTGGCCA	359
Qy	299	TTCTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTG	358
Db	360	TCTTGGTCTGTGGCTGCTCACCTGGAGGGAACCTGTCCAATGTCTTCAGTCTGGCCATGA	419
Qy	359	ATGGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAA	418
Db	420	AGGGGGACATGAACCTCAGCATTGTGATGACCACCTGCTCCACCTTCTGTGCCCTTGGA	479
Qy	419	TGATGCCACTCTGCATTTATCTCTACACC---TGGTCCTGGAGTCTTCAGCAGAATCTCA	475
Db	480	TGATGCCTCTCCTCCTGTACATCTACTCCAGGGGGATCTATGATGGGGACCTGAAGGACA	539
Qy	476	CCATTCCCTTATCAGAACATAGGAATTACCCTTGTGTGCCTGACCATTCCCTGTGGCCTTTG	535
Db	540	AGGTGCCCTATAAAGGCATCGTGATATCACTGGTCCTGGTTCTCATTCCTTGCACCATAG	599
Qy	536	GTGTCTATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATTGGGGCCG	595
Db	600	GGATCGTCCTCAAATCCAAACGGCCACAATACATGCGCTATGTATCAAGGGAGGGATGA	659
Qy	596	TTGTTGGTGGGGTCCTCCTTCTGGTGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGAT	655

Db	660	TCATCATCTCTTGTGCAGTGTGGCCGTACAGTTCTCTCTGCCATCAATGTGGGGAAGA	719
Qy	656	CTTGGAATTCAGACATCAC-----CCTTCTGACCATCAGTTTCATCTTTCCCTTTGATTG	709
Db	720	GCATCATGTTTGCCATGACACCACTCTTGATTGCCACCTCCTCCCTGATGCCTTTTATTG	779
Qy	710	GCCATGTCACGGGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGCCAAAGGTGCAGGA	769
Db	780	GCTTCTGCTGGGTTATGTTCTCTCTGCTCTTCTGCCTCAATGGACGGTGCAGACGCA	839
Qy	770	CAATTCCTTAGAAACTGGAGCTCAGAAATATTCAGATGTGCATCACCATGCTCCAGTTAT	829
Db	840	CTGTCAGCATGGAGACTGGATGCCAAAATGTCCAACCTCTGTTCCACCATCCTCAATGTGG	899
Qy	830	CTTTCAC TGTGAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTCC	889
Db	900	CCTTTCACCTGAAGTCATTGGACCACTTTTCTTCTTTCCCTCCTCTACATGATTTTCC	959
Qy	890	AGCTGATAGATGGATTTCTTATTGTTGCAGCATATCAGACGTACAAGAGGAGATTGAAGA	949
Db	960	AGCTTGGAGAAGGGCTTCTCCTCATTGCCATATTTTGGTGTCTATGAGAAATCAAGACTC	1019
Qy	950	ACAAACATGGAAAAAAGA	967
Db	1020	CCAAGGATAAAACAAAAA	1037

Search completed: March 25, 2004, 16:38:33
Job time : 592 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2004, 16:01:20 ; Search time 122 Seconds
(without alignments)
5158.314 Million cell updates/sec

Title: US-10-091-628-1
Perfect score: 1134
Sequence: 1 atgagagccaattgttccag.....acatcacttcatgtgaatag 1134

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	% Match	Query Length	DB ID	Description	
	1	320.4	28.3	2263	1	US-08-176-126B-1	Sequence 1, Appli
	2	320.4	28.3	2263	2	US-08-669-435-1	Sequence 1, Appli
	3	320.4	28.3	2263	5	PCT-US94-14431A-1	Sequence 1, Appli
	4	297.8	26.3	1047	1	US-08-176-126B-3	Sequence 3, Appli
	5	297.8	26.3	1047	2	US-08-669-435-3	Sequence 3, Appli
	6	297.8	26.3	1047	5	PCT-US94-14431A-3	Sequence 3, Appli
c	7	77.8	6.9	310	4	US-09-833-381-317	Sequence 317, App
	8	64	5.6	927	4	US-09-252-991A-7387	Sequence 7387, Ap
	9	64	5.6	978	4	US-09-252-991A-7319	Sequence 7319, Ap
c	10	64	5.6	1008	4	US-09-252-991A-7300	Sequence 7300, Ap
	11	58.6	5.2	374	4	US-09-833-381-318	Sequence 318, App

	12	58.2	5.1	972	4	US-09-540-236-963	Sequence 963, App
c	13	58.2	5.1	269223	4	US-09-596-002-41	Sequence 41, Appl
c	14	55.8	4.9	891	4	US-09-252-991A-1376	Sequence 1376, Ap
	15	55.8	4.9	948	4	US-09-252-991A-1144	Sequence 1144, Ap
	16	55.8	4.9	1050	4	US-09-489-039A-1413	Sequence 1413, Ap
	17	55.8	4.9	1077	4	US-09-252-991A-1231	Sequence 1231, Ap
c	18	55.8	4.9	1719	4	US-09-252-991A-1327	Sequence 1327, Ap
	19	51.2	4.5	7218	1	US-08-232-463-14	Sequence 14, Appl
c	20	39	3.4	592	4	US-09-442-143A-5	Sequence 5, Appli
c	21	39	3.4	5403	4	US-09-442-143A-3	Sequence 3, Appli
	22	35.2	3.1	505	4	US-09-621-976-15639	Sequence 15639, A
c	23	34.6	3.1	289	3	US-09-007-005-17	Sequence 17, Appl
c	24	34.6	3.1	289	3	US-09-244-796-17	Sequence 17, Appl
	25	34.6	3.1	978	4	US-09-328-352-974	Sequence 974, App
c	26	34.6	3.1	1866	4	US-09-543-681A-2572	Sequence 2572, Ap
c	27	34.2	3.0	66804	4	US-09-740-041-3	Sequence 3, Appli
c	28	34	3.0	615	4	US-09-489-039A-303	Sequence 303, App
c	29	34	3.0	1242	4	US-09-489-039A-391	Sequence 391, App
c	30	33.6	3.0	505	4	US-09-621-976-15639	Sequence 15639, A
c	31	33.2	2.9	1911	4	US-09-833-381-1045	Sequence 1045, Ap
c	32	33	2.9	193303	4	US-09-497-855A-37	Sequence 37, Appl
c	33	33	2.9	193303	4	US-09-497-855A-44	Sequence 44, Appl
c	34	32.8	2.9	423	4	US-09-252-991A-14207	Sequence 14207, A
c	35	32.8	2.9	474	4	US-09-621-976-18033	Sequence 18033, A
	36	32.8	2.9	2151	4	US-09-252-991A-14112	Sequence 14112, A
c	37	32.8	2.9	2223	4	US-09-252-991A-14116	Sequence 14116, A
	38	32.8	2.9	3104	4	US-09-418-780A-2	Sequence 2, Appli
	39	32.8	2.9	3106	4	US-09-392-714-4	Sequence 4, Appli
	40	32.8	2.9	3159	4	US-09-252-991A-14020	Sequence 14020, A
c	41	32.2	2.8	7218	1	US-08-232-463-14	Sequence 14, Appl
	42	32	2.8	1776	4	US-09-679-686B-11	Sequence 11, Appl
c	43	31.8	2.8	399	4	US-09-621-976-8976	Sequence 8976, Ap
c	44	31.8	2.8	861	4	US-09-252-991A-940	Sequence 940, App
	45	31.8	2.8	1068	4	US-09-252-991A-1105	Sequence 1105, Ap

ALIGNMENTS

RESULT 1

US-08-176-126B-1

; Sequence 1, Application US/08176126B

; Patent No. 5589358

; GENERAL INFORMATION:

; APPLICANT: Dawson, Paul A.

; TITLE OF INVENTION: ILEAL BILE ACID TRANSPORTER COMPOSITIONS AND

; TITLE OF INVENTION: METHODS

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White & Durkee

; STREET: P.O. Box 4433

; CITY: Houston

; STATE: Texas

; COUNTRY: US

; ZIP: 77210

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

```

;   COMPUTER:  IBM PC compatible
;   OPERATING SYSTEM:  PC-DOS/MS-DOS/ASCII
;   SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/08/176,126B
;   FILING DATE:  29-DEC-1993
;   CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;   NAME:  Parker, David L.
;   REGISTRATION NUMBER:  32,165
;   REFERENCE/DOCKET NUMBER:  WAKE:002/PAR
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  (512) 418-3000
;   TELEFAX:  (512) 474-7577
;   TELEX:  na
;   INFORMATION FOR SEQ ID NO:  1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH:  2263 base pairs
;   TYPE:  nucleic acid
;   STRANDEDNESS:  single
;   TOPOLOGY:  linear
;   MOLECULE TYPE:  cDNA
;   FEATURE:
;   NAME/KEY:  CDS
;   LOCATION:  109..1152
US-08-176-126B-1

```

```

Query Match          28.3%;  Score 320.4;  DB 1;  Length 2263;
Best Local Similarity  60.8%;  Pred. No. 8.2e-91;
Matches 522;  Conservative  0;  Mismatches 336;  Indels  0;  Gaps  0;

```

```

Qy      80 ATGGAAACCTGGAGCTCGTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGT 139
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     188 ACGCCATCCTCAGCGTGGTGTGAGCACCGTGCTCACAATCCTCCTAGCCTTGGTGATGT 247

Qy     140 TCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTGCACATCAGGAGACCCCTGGG 199
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     248 TTTCCATGGGGTGCAATGTGGAAGTCCACAAGTTTCTGGGACACCTAAGGCGGCCATGGG 307

Qy     200 GCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGG 259
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     308 GCATCGTCGTGGGCTTCCTCTGTCTAGTTTGAATCATGCCTCTCACAGGTTTCGTCTGT 367

Qy     260 CCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCC 319
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     368 CCGTGGCCTTTTGGCATCCTCCAGTGCAAGCTGTGGTGGTGTGCTGATCCAGGTTGCTGCC 427

Qy     320 CGGGGGGACCATCTCTAACATTTTTCACCTTCTGGGTTGATGGAGATATGGATCTCAGCA 379
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     428 CTGGAGGAAGTGCCTCCAATATCCTAGCCTATTGGGTAGATGGCGACATGGACCTCAGCG 487

Qy     380 TCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATC 439
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     488 TTAGCATGACCACCTGCTCCACGCTGCTTGCCCTTGAATGATGCCCCTTTGCCTCTTCA 547

Qy     440 TCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATTCTTATCAGAACATAGGAA 499
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 548 TCTATACCAAGATGTGGGTTGACTCAGGGACGATTGTGATTCCCTTATGACAGCATTGGCA 607
 Qy 500 TTACCCCTTGTTGCTGCCTGACCATTCCCTGTGGCCTTTGGTGTCTATGTGAATTACAGATGGC 559
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 608 CTTCTCTGGTTGCTCTTGTATTCTCTGTTTCCATTGGAATGTATGTGAATCACAAATGGC 667
 Qy 560 CAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTTGGTGGGGTCCCTCCTTCTGG 619
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 668 CCCAAAAGCAAAGATCATACTTAAAATTGGATCCATCGCAGGTGCAATTCTCATTGTTC 727
 Qy 620 TGGTCGCAGTTGCTGGTGTGGTCCCTGGCGAAAGGATCTTGGAATTCAGACATCACCCCTTC 679
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 728 TCATCGCTGTGGTTGGAGGAATACTGTACCAAAGTGCCTGGACCATTGAACCCAAGCTGT 787
 Qy 680 TGACCATCAGTTTCATCTTTCCCTTTGATTGGCCATGTCACGGGTTTCTGCTGGCACTTT 739
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 788 GGATTATAGGAACCATATATCCTATAGCTGGCTACGGCCTGGGGTTTTCTGGCTAGAA 847
 Qy 740 TTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCCCTTAGAAACTGGAGCTCAGAATA 799
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 848 TTGCTGGTCAACCCTGGTACAGGTGCCGAACAGTTGCCTTGAAACCGGGTTGCAGAACA 907
 Qy 800 TTCAGATGTGCATCACCATGCTCCAGTTATCTTCACTGCTGAGCACTTGGTCCAGATGT 859
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 908 CTCAGCTGTGTTCCACCATTGTGCAGCTTTCCTTCAGCCCTGAGGACCTCAACCTTGTGT 967
 Qy 860 TGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGATGGATTCTTATTGTTGCAG 919
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 968 TCACCTTCCCCCTCATCTACAGCATCTTCCAGATCGCCTTTGCAGCAATACTATTAGGAG 1027
 Qy 920 CATATCAGACGTACAAGA 937
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 1028 CTTATGTGCGCATACAAGA 1045

RESULT 2

US-08-669-435-1

; Sequence 1, Application US/08669435

; Patent No. 5869265

; GENERAL INFORMATION:

; APPLICANT: Dawson, Paul A.

; TITLE OF INVENTION: ILEAL BILE ACID TRANSPORTER COMPOSITIONS AND

; TITLE OF INVENTION: METHODS

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White & Durkee

; STREET: P.O. Box 4433

; CITY: Houston

; STATE: Texas

; COUNTRY: US

; ZIP: 77210

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

```

; APPLICATION NUMBER: US/08/669,435
; FILING DATE: 26-JUN-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/176,126
; FILING DATE: 29-DEC-1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: WAKE:002/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; TELEX: na
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2263 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 109..1152
US-08-669-435-1

```

```

Query Match          28.3%; Score 320.4; DB 2; Length 2263;
Best Local Similarity 60.8%; Pred. No. 8.2e-91;
Matches 522; Conservative 0; Mismatches 336; Indels 0; Gaps 0;

```

```

Qy      80 ATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGT 139
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      188 ACGCCATCCTCAGCGTGGTGATGAGCACC GTGCTCACAATCCTCCTAGCCTTGGTGATGT 247

Qy      140 TCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCGCACATCAGGAGACCCTGGG 199
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      248 TTTCCATGGGGTGCAATGTGGAAGTCCACAAGTTTCTGGGACACCTAAGCGGCCATGGG 307

Qy      200 GCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGG 259
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      308 GCATCGTCGTGGGCTTCCTCTGTCAAGTTTGAATCATGCCTCTCACAGTTTCGTCCTGT 367

Qy      260 CCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCC 319
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      368 CCGTGGCCTTTGGCATCCTCCAGTGCAAGCTGTGGTGGTGCTGATCCAGGGTTGCTGCC 427

Qy      320 CGGGGGGACCATCTCTAACAATTTTACCTTCTGGGTTGATGGAGATATGGATCTCAGCA 379
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      428 CTGGAGGAACTGCCTCCAATATCCTAGCCTATTGGGTAGATGGCGACATGGACCTCAGCG 487

Qy      380 TCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATC 439
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      488 TTAGCATGACCACCTGCTCCACGCTGCTTGCCCTTGAATGATGCCCTTTGCTCTTCA 547

Qy      440 TCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATTCTTATCAGAACATAGGAA 499
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db	548	TCTATACCAAGATGTGGGTTGACTCAGGGACGATTGTGATTTCCTTATGACAGCATTGGCA	607
Qy	500	TTACCCCTTGTGTGCCTGACCATTCTGTGGCCTTTGGTGTCTATGTGAATTACAGATGGC	559
Db	608	CTTCTCTGGTTGCTCTTGTATTCTGTTCATTGGAATGTATGTGAATCACAAATGGC	667
Qy	560	CAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTTGGTGGGGTCCTCCTTCTGG	619
Db	668	CCCAAAAAGCAAAGATCATACTTAAATTTGGATCCATCGCAGGTGCAATTCTCATTGTTC	727
Qy	620	TGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGATCTTGAATTCAGACATCACCCCTTC	679
Db	728	TCATCGCTGTGGTTGGAGGAATACTGTACCAAAGTGCCTGGACCATTGAACCCAAGCTGT	787
Qy	680	TGACCATCAGTTTCATCTTTTCCTTTGATTGGCCATGTCACGGGTTTTCTGCTGGCACTTT	739
Db	788	GGATTATAGGAACCATATATCCTATAGCTGGCTACGGCCTGGGGTTTTTCCTGGCTAGAA	847
Qy	740	TTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCCCTTAGAAACTGGAGCTCAGAATA	799
Db	848	TTGCTGGTCAACCCTGGTACAGGTGCCGAACAGTTGCCTTGGAACCGGGTTGCAGAACA	907
Qy	800	TTCAGATGTGCATCACCATGCTCCAGTTATCTTTCACTGCTGAGCACTTGGTCCAGATGT	859
Db	908	CTCAGCTGTGTCCACCATTGTGCAGCTTTCCTTCAGCCCTGAGGACCTCAACCTTGTGT	967
Qy	860	TGAGTTTCCCACTGGCCTATGGACTCTTCAGCTGATAGATGGATTTCTTATTGTTGCAG	919
Db	968	TCACCTTCCCCCTCATCTACAGCATCTTCAGATCGCCTTTGCAGCAATACTATTAGGAG	1027
Qy	920	CATATCAGACGTACAAGA	937
Db	1028	CTTATGTGCGCATACAAGA	1045

PCT-US94-14431A-1

; GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: ILEAL BILE ACID TRANSPORTER COMPOSITIONS

```
; NUMBER OF SEQUENCES: 11
```

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee

STREET: P. O. Box 4433

CITY: Houston

STATE: Texas

; COUNTRY: United States of America

ZIP: 77210

; COMPUTER READABLE FORM:

```
; MEDIUM TYPE: Floppy disk
```

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII

```

;      SOFTWARE:  PatentIn Release #1.0, Version

```

```

;      SOFTWARE:  #1.25

```

; CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/14431A

```

; FILING DATE: 29-DEC-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: USSN 08/176,126
; FILING DATE: 29-DEC-1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: PARKER, DAVID L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: WAKE005P--
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924(1) GENERAL INFORMATION:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2263 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 109..1152
PCT-US94-14431A-1

```

```

Query Match          28.3%; Score 320.4; DB 5; Length 2263;
Best Local Similarity 60.8%; Pred. No. 8.2e-91;
Matches 522; Conservative 0; Mismatches 336; Indels 0; Gaps 0;

```

```

Qy      80 ATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGT 139
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     188 ACGCCATCCTCAGCGTGGTGATGAGCACC GTGCTCACAATCCTCCTAGCCTTGGTGATGT 247

Qy     140 TCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTTCGCACATCAGGAGACCCTGGG 199
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     248 TTTCCATGGGGTGCAATGTGGAAGTCCACAAGTTTCTGGGACACCTAAGGCGGCCATGGG 307

Qy     200 GCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGG 259
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     308 GCATCGTCGTGGGCTTCTCTGTGAGTTTGAATCATGCCTCTCACAGGTTTCGTCTGT 367

Qy     260 CCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCC 319
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     368 CCGTGGCCCTTTGGCATCCTCCCAGTGCAAGCTGTGGTGGTGCTGATCCAGGGTTGCTGCC 427

Qy     320 CGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGATGGAGATATGGATCTCAGCA 379
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     428 CTGGAGGAAGTGCCTCCAATATCCTAGCCTATTGGGTAGATGGCGACATGGACCTCAGCG 487

Qy     380 TCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATC 439
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     488 TTAGCATGACCACCTGCTCCACGCTGCTTGCCCTTGAATGATGCCCTTTGCCTCTTCA 547

Qy     440 TCTACACCTGGTCTGAGTCTTCAGCAGAATCTCACCATTCCCTTATCAGAACATAGGAA 499
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     548 TCTATACCAAGATGTGGGTTGACTCAGGGACGATTGTGATTCTTATGACAGCATTGGCA 607

```


QY 500 TTACCCCTTGTGTGCCTGACCAATCC'TGTGGCCTTTGGTGTCTATGTGAATTACAGATGGC 559
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 608 CTTCTCTGGTTGCTCTTGTATTCCCTGTTTCCATTGGAATGTATGTGAATCACAAATGGC 667
 QY 560 CAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTTGGTGGGGTCCTCCTTCTGG 619
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 668 CCCAAAAGCAAAGATCATACTTAAAATTGGATCCATCGCAGGTGCAATTCTCATTGTTC 727
 QY 620 TGGTCGCAGTTGCTGGTGTGGTCC'TGGCGAAAGGATCTTGAATTCAGACATCACCCCTTC 679
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 728 TCATCGCTGTGGTTGGAGGAATACTGTACCAAAGTGCCTGGACCATTGAACCCAAGCTGT 787
 QY 680 TGACCATCAGTTTCATCTTTCCTTTGATTGGCCATGTCACGGGTTTTCTGCTGGCACTTT 739
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 788 GGATTATAGGAACCATATATCCTATAGCTGGCTACGGCCTGGGGTTTTCTGCTAGAA 847
 QY 740 TTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCCCTTAGAAACTGGAGCTCAGAATA 799
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 848 TTGCTGGTCAACCCTGGTACAGGTGCCGAACAGTTGCCTTGGAAACCGGGTTGCAGAACA 907
 QY 800 TTCAGATGTGCATCACCATGCTCCAGTTATCTTCACTGCTGAGCACTTGGTCCAGATGT 859
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 908 CTCAGCTGTGTTCCACCATTGTGCAGCTTTCCTTCAGCCCTGAGGACCTCAACCTTGTGT 967
 QY 860 TGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGATGGATTTCTTATTGTTGCAG 919
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 968 TCACCTTCCCCCTCATCTACAGCATCTTCCAGATCGCCTTTGCAGCAATACTATTAGGAG 1027
 QY 920 CATATCAGACGTACAAGA 937
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 1028 CTTATGTCGCATACAAGA 1045

RESULT 4

US-08-176-126B-3

; Sequence 3, Application US/08176126B

; Patent No. 5589358

; GENERAL INFORMATION:

; APPLICANT: Dawson, Paul A.

; TITLE OF INVENTION: ILEAL BILE ACID TRANSPORTER COMPOSITIONS AND

; TITLE OF INVENTION: METHODS

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White & Durkee

; STREET: P.O. Box 4433

; CITY: Houston

; STATE: Texas

; COUNTRY: US

; ZIP: 77210

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/176,126B

```

; FILING DATE: 29-DEC-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: WAKE:002/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; TELEX: na
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1047 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1044
US-08-176-126B-3

```

```

Query Match          26.3%; Score 297.8; DB 1; Length 1047;
Best Local Similarity 58.5%; Pred. No. 7.2e-84;
Matches 518; Conservative 0; Mismatches 367; Indels 0; Gaps 0;

```

```

Qy      80 ATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGT 139
      || | ||| | | | | | | | | | | | | | | | | | | | | | | | |
Db      80 ATAACATCCTAAGTGTGGTCCCTAAGTACGGTGTGCTGACCATCCTGTTGGCCTTGCTGATGT 139

Qy     140 TCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTTCGCACATCAGGAGACCCTGGG 199
      ||| | ||||| | | | | | | | | | | | | | | | | | | | | | | |
Db     140 TCTCCATGGGATGCAACGTGGAAATCAAGAAATTTCTAGGGCACATAAAGCGGCCGTGGG 199

Qy     200 GCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGG 259
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     200 GCATTTGTGTGGCTTCCTCTGTTCAGTTTGAATCATGCCCTCACAGGATTCATCCTGT 259

Qy     260 CCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCC 319
      | | |||| | | | | | | | | | | | | | | | | | | | | | | | |
Db     260 CGGTGGCCTTTGACATCCTCCCGCTCCAGGCCGTAGTGGTGTCTATTATAGGATGCTGCC 319

Qy     320 CGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGATGGAGATATGGATCTCAGCA 379
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     320 CTGGAGGAAGTGCCTCCAATATCTTGGCCTATTGGGTGATGGCGACATGGACCTGAGCG 379

Qy     380 TCAGTATGACAACTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATC 439
      |||| | |||| | | | | | | | | | | | | | | | | | | | | | |
Db     380 TCAGCATGACCACATGCTCCACACTGCTTGCCCTCGGAATGATGCCGCTGTGCCTCCTTA 439

Qy     440 TCTACACCTGGTCTCGAGTCTTCAGCAGAATCTCACCATTTCCTTATCAGAACATAGGAA 499
      |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     440 TCTATACCAAATGTGGGTCGACTCTGGGAGCATCGTAATCCCTATGATAACATAGGTA 499

Qy     500 TTACCCTTGCTGCTGCTGACCATTCCTGTGGCCTTTGGTGTCTATGTGAATTACAGATGGC 559
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     500 CATCTCTGGTTGCTCTCGTTGTTCTGTTTCCATTGGAATGTTTGTTAATCACAAATGGC 559

```

Qy 560 CAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTTGGTGGGGTCCTCCTTCTGG 619
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 560 CCCAAAAAGCAAAGATCATACTTAAATTTGGGTCCATCGCGGGCGCCATCCTCATTGTGC 619
 Qy 620 TGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGATCTTGAATTCAGACATCACCTTC 679
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 620 TCATAGCTGTGGTTGGAGGAATATTGTACCAAAGCGCCTGGATCATTGCTCCCAAACGT 679
 Qy 680 TGACCATCAGTTTCATCTTTCCTTTGATTGGCCATGTCACGGGTTTTCTGCTGGCACTTT 739
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 680 GGATTATAGGAACAATATTTCTGTGGCGGGTTACTCCCTGGGGTTTCTTCTGGCTAGAA 739
 Qy 740 TTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCTTAGAAACTGGAGCTCAGAATA 799
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 740 TTGCTGGTCTACCCTGGTACAGGTGCCGAACGGTTGCTTTTGAAACGGGGATGCAGAACA 799
 Qy 800 TTCAGATGTGCATCACCATGCTCCAGTTATCTTTCAGTGTGAGCACTTGGTCCAGATGT 859
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 800 CGCAGCTATGTTCCACCATCGTTCAGCTCTCCTTCACTCCTGAGGAGCTCAATGTCGTAT 859
 Qy 860 TGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGATGGATTTCTTATTGTTGCAG 919
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 860 TCACCTTCCCGCTCATCTACAGCATTTTCCAGCTCGCCTTTGCCGCAATATTCTTAGGAT 919
 Qy 920 CATATCAGACGTACAAGAGGAGATTGAAGAACAAACATGGAAAAA 964
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 920 TTTATGTGGCATAACAAGAAATGTCATGGAAAAAACAAGGCAGAAA 964

RESULT 5

US-08-669-435-3

; Sequence 3, Application US/08669435

; Patent No. 5869265

; GENERAL INFORMATION:

; APPLICANT: Dawson, Paul A.

; TITLE OF INVENTION: ILEAL BILE ACID TRANSPORTER COMPOSITIONS AND

; TITLE OF INVENTION: METHODS

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White & Durkee

; STREET: P.O. Box 4433

; CITY: Houston

; STATE: Texas

; COUNTRY: US

; ZIP: 77210

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/669,435

; FILING DATE: 26-JUN-1996

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/176,126

```

; FILING DATE: 29-DEC-1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: WAKE:002/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; TELEX: na
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1047 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1044
US-08-669-435-3

```

```

Query Match          26.3%; Score 297.8; DB 2; Length 1047;
Best Local Similarity 58.5%; Pred. No. 7.2e-84;
Matches 518; Conservative 0; Mismatches 367; Indels 0; Gaps 0;

```

```

Qy      80 ATGGAAACCTGGAGCTCGTTTTACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGT 139
      || || || | || | || | || | || | || | || | || | || | || |
Db      80 ATAACATCCTAAGTGTGGTCCCTAAGTACGGTGTGACCATCCTGTTGGCCTTGGTGATGT 139

Qy     140 TCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTTCGCACATCAGGAGACCCTGGG 199
      || || || || || || || || || || || || || || || || || || || ||
Db     140 TCTCCATGGGATGCAACGTGGAAATCAAGAAATTTCTAGGGCACATAAAGCGGCCGTGGG 199

Qy     200 GCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGG 259
      || || || || || || || || || || || || || || || || || || || ||
Db     200 GCATTGTGTGGCTTCCTCTGTCTAGTTTGGGAATCATGCCCTCACAGGATTCATCCTGT 259

Qy     260 CCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCC 319
      | | || || | || || || || || || || || || || || || || || || ||
Db     260 CGGTGGCCTTTGACATCCTCCCGCTCCAGGCCGTAGTGGTGTCTATTATAGGATGCTGCC 319

Qy     320 CGGGGGGCACCATCTCTAACATTTTACCTTCTGGGTTGATGGAGATATGGATCTCAGCA 379
      | || || || || || || || || || || || || || || || || || || || ||
Db     320 CTGGAGGAACTGCCTCCAATATCTTGGCCTATTGGGTCGATGGCGACATGGACCTGAGCG 379

Qy     380 TCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATC 439
      || || || || || || || || || || || || || || || || || || || ||
Db     380 TCAGCATGACCACATGCTCCACACTGCTTGCCCTCGGAATGATGCCGCTGTGCCTCCTTA 439

Qy     440 TCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATTCCCTTATCAGAACATAGGAA 499
      || || || || || || || || || || || || || || || || || || || ||
Db     440 TCTATACAAAATGTGGGTCGACTCTGGGAGCATCGTAATCCCTATGATAACATAGGTA 499

Qy     500 TTACCCTTGTGTGCCTGACCATTCTGTGGCCTTTGGTGTCTATGTGAATTACAGATGGC 559
      | || || || || || || || || || || || || || || || || || || || ||
Db     500 CATCTCTGGTTGCTCTCGTTGTTCTGTTTCCATTGGAATGTTTGTAAATCACAAATGGC 559

```

Qy 560 CAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTTGGTGGGGTCCCTCCTTCTGG 619
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 560 CCCAAAAAGCAAAGATCATACTTAAAATTGGGTCCATCGCGGGCGCCATCCTCATTGTGC 619
 Qy 620 TGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGATCTTGGAATTCAGACATCACCCCTTC 679
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 620 TCATAGCTGTGGTTGGAGGAATATTGTACCAAAGCGCCTGGATCATTGCTCCCAAAGTGT 679
 Qy 680 TGACCATCAGTTTCATCTTTCTTTGATTGGCCATGTCACGGGTTTTCTGCTGGCACTTT 739
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 680 GGATTATAGGAACAATATTTCTGTGGCGGGTTACTCCCTGGGGTTTCTTCTGGCTAGAA 739
 Qy 740 TTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCTTAGAAACTGGAGCTCAGAATA 799
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 740 TTGCTGGTCTACCCTGGTACAGGTGCCGAACGGTTGCTTTTGAAACGGGGATGCAGAACA 799
 Qy 800 TTCAGATGTGCATCACCATGCTCCAGTTATCTTTCACTGCTGAGCACTTGGTCCAGATGT 859
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 800 CGCAGCTATGTTCCACCATCGTTAGCTCTCCTTCACTCCTGAGGAGCTCAATGTCGTAT 859
 Qy 860 TGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGATGGATTTCTTATTGTTGCAG 919
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 860 TCACCTTCCCGCTCATCTACAGCATTTTCCAGCTCGCCTTTGCCGCAATATTCTTAGGAT 919
 Qy 920 CATATCAGACGTACAAGAGGAGATTGAAGAACAAACATGGAAAAA 964
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 920 TTTATGTGGCATAACAAGAAATGTCATGGAAAAACAAGGCAGAAA 964

RESULT 6

PCT-US94-14431A-3

; Sequence 3, Application PC/TUS9414431A

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: ILEAL BILE ACID TRANSPORTER COMPOSITIONS

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White & Durkee

; STREET: P. O. Box 4433

; CITY: Houston

; STATE: Texas

; COUNTRY: United States of America

; ZIP: 77210

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII

; SOFTWARE: PatentIn Release #1.0, Version

; SOFTWARE: #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US94/14431A

; FILING DATE: 29-DEC-1994

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: USSN 08/176,126

; FILING DATE: 29-DEC-1993

```

; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: PARKER, DAVID L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: WAKE005P--
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924(1) GENERAL INFORMATION:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1047 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1044
PCT-US94-14431A-3

```

Qy	80	ATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGT	139
Db	80	ATAACATCCTAAGTGTGGTCCTAAGTACGGTGCTGACCATCCTGTTGGCCTTGGTGATGT	139
Qy	140	TCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCGCACATCAGGAGACCCTGGG	199
Db	140	TCTCCATGGGATGCAACGTGGAAATCAAGAAATTTCTAGGGCACATAAAGCGGCCGTGGG	199
Qy	200	GCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGG	259
Db	200	GCATTGTGTGTGGCTTCCTCTGTCAAGTTTGGAAATCATGCCCTCACAGGATTCATCCTGT	259
Qy	260	CCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCC	319
Db	260	CGGTGGCCTTTGACATCCTCCCGCTCCAGGCCGTAGTGGTGCTCATTATAGGATGCTGCC	319
Qy	320	CGGGGGGCACCATCTCTAACATTTTACCTTCTGGGTTGATGGAGATATGGATCTCAGCA	379
Db	320	CTGGAGGAAGTGCCTCCAATATCTTGGCCTATTGGGTCGATGGCGACATGGACCTGAGCG	379
Qy	380	TCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATC	439
Db	380	TCAGCATGACCACATGCTCCACACTGCTTGCCCTCGGAATGATGCCGCTGTGCCTCCTTA	439
Qy	440	TCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATTCCCTTATCAGAACATAGGAA	499
Db	440	TCTATACCAAATGTGGGTCGACTCTGGGAGCATCGTAATTCCTTATGATAACATAGGTA	499
Qy	500	TTACCCTTGTGTGCCTGACCATTCCCTGTGGCCTTTGGTGTCTATGTGAATTACAGATGGC	559
Db	500	CATCTCTGGTTGCTCTCGTTGTTTCTGTTTCCATTGGAATGTTTGTTAATCACAAATGGC	559

Qy 748 CAGTCTTGGCAAAGGTGCAGGACAATTTCTTAGAACTGGAGCTCAGAATATTCAGATG 807
 || | ||| | ||||| | ||| | |||| | ||||| | |||| | ||
 Db 221 CAACCCTGGTACAGGTGCCGAACAGTAGCCTTGGAACCTGGAATGCAGAACTCAGCTG 162

Qy 808 TGCATCACCATGCTCCAGTTATCTTTCTACTGCTGAGCACTTGGTCCAGATGTTGAGTTTC 867
 ||| ||||| | ||| | |||| | |||| | ||| | ||| ||||| | |||
 Db 161 TGCTCCACCATTGTACAGCTCTCCTTCTCCCCGAGGATCTCAACCTGGTGTTCACCTTC 102

Qy 868 CCACTGGCCTATGGACTCTTCCAGCTGATAGATGGATTCTTATTGTTGCAGCATATCAG 927
 ||||| |||| | ||||| || | || | || | || | ||
 Db 101 CCACTCATCTATACTGTTTTCCAGCTCGTCTTTGCAGCAGTNATATTAGGNATTTATGTC 42

Qy 928 ACGTACAAGAGGAGATTGAAGAACAAACATG 958
 || |||| || | | || || || ||
 Db 41 ACATACAGGAAATGTTATGGAAAAAATGATG 11

RESULT 8

US-09-252-991A-7387

; Sequence 7387, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 7387

; LENGTH: 927

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-7387

Query Match 5.6%; Score 64; DB 4; Length 927;

Best Local Similarity 48.9%; Pred. No. 6e-10;

Matches 172; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

Qy 91 GAGCTCGTTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGTTCTCTTTGGGA 150
 || | || | || || || | || | |||| | ||||| || | ||||
 Db 43 GATCCCATCCTGACCCTGTTCTCTCCCATCGCACTGGGCATCATCATGCTCGGTCTCGGA 102

Qy 151 TGTTCCGTGGAGATCCGGAAGCTGTGGTGCACATCAGGAGACCCTGGGGCATTGCTGTG 210
 ||| || | | | || | || | || | || | ||
 Db 103 CTGTCCCTGACCCCGGCCGACTTCTCCGCGTGGTGCCTACCCGAAGCCGGTGCTGGTC 162

Qy 211 GGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGGCCATTAGCTTT 270
 || ||| | ||||| | | || |||| | | || || | || | ||||
 Db 163 GGCTTGGTGTGCCAGATCGTCCTGCTGCCCCCTGGCCTGTTTCTGATCGTCCAGGGCTTC 222

Qy 271 TCTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCCCCGGGGGGCACC 330


```

      | | | | | | | | | | | | | | | | | | | | | | | |
Db      223 GCCCTGGAGGCGGCCCTGGCGGTGGGCATGATGTTGCTGGCGGCCTCGCCCGGCGGCACC 282

Qy      331 ATCTCTAACATTTTCACCTTCTGGGTTGATGGAGATATGGATCTCAGCATCAGTATGACA 390
      | | | | | | | | | | | | | | | | | | | | | | | |
Db      283 ACCGCCAACCTCTACAGCCACCTGGCGCATGGCGACGTGGCACTGAACATCACCTGACC 342

Qy      391 ACCTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATCTCT 442
      | | | | | | | | | | | | | | | | | | | | | |
Db      343 GCGGTGAACTCGGTGATCGCCATCCTCACCATGCCGCTGATCGTCAATCTGT 394

```

RESULT 9

US-09-252-991A-7319

; Sequence 7319, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 7319

; LENGTH: 978

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-7319

```

Query Match          5.6%;  Score 64;  DB 4;  Length 978;
Best Local Similarity 48.9%;  Pred. No. 6.2e-10;
Matches 172;  Conservative 0;  Mismatches 180;  Indels 0;  Gaps 0;

```

```

Qy      91 GAGCTCGTTTTACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGTTCTCTTTGGGA 150
      | | | | | | | | | | | | | | | | | | | | | |
Db      68 GATCCCATCCTGACCCCTGTTCTCTCCCATCGCACTGGGCATCATCATGCTCGGTCTCGGA 127

Qy     151 TGTTCCGTGGAGATCCGGAAGCTGTGGTGCACATCAGGAGACCCTGGGGCATTGCTGTG 210
      | | | | | | | | | | | | | | | | | | | | |
Db     128 CTGTCCCTGACCCCGGCCACTTCTCCGCGTGGTGCCTACCCGAAGCCGGTGCTGGTC 187

Qy     211 GGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGGCCATTAGCTTT 270
      | | | | | | | | | | | | | | | | | | | | |
Db     188 GGCCTGGTGTGCCAGATCGTCCTGTGCCCCCTGGCCTGTTTCTGATCGTCCAGGGCTTC 247

Qy     271 TCTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCCCCGGGGGCACC 330
      | | | | | | | | | | | | | | | | | | | | |
Db     248 GCCCTGGAGGCGGCCCTGGCGGTGGGCATGATGTTGCTGGCGGCCTCGCCCGGCGGCACC 307

Qy     331 ATCTCTAACATTTTCACCTTCTGGGTTGATGGAGATATGGATCTCAGCATCAGTATGACA 390
      | | | | | | | | | | | | | | | | | | | | |

```

Db 308 ACCGCCAACCTCTACAGCCACCTGGCGCATGGCGACGTGGCACTGAACATCACCTTGACC 367

Qy 391 ACCTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATCTCT 442

Db 368 GCGGTGAACTCGGTGATCGCCATCCTCACCATGCCGCTGATCGTCAATCTGT 419

RESULT 10

US-09-252-991A-7300/c

; Sequence 7300, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 7300

; LENGTH: 1008

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-7300

Query Match 5.6%; Score 64; DB 4; Length 1008;

Best Local Similarity 48.9%; Pred. No. 6.3e-10;

Matches 172; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

Qy 91 GAGCTCGTTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGTTCTCTTTGGGA 150

Db 891 GATCCCATCCTGACCCTGTTCTCCCCATCGCACTGGGCATCATCATGCTCGGTCTCGGA 832

Qy 151 TGTTCCGTGGAGATCCGGAAGCTGTGGTTCGCACATCAGGAGACCCTGGGGCATTGCTGTG 210

Db 831 CTGTCCCTGACCCCGGCCACTTCCTCCGCGTGGTGGCTACCCGAAGCCGGTGTGGTC 772

Qy 211 GGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGGCCATTAGCTTT 270

Db 771 GGCCTGGTGTGCCAGATCGTCCTGCTGCCCCTGGCCTGTTTCTGATCGTCCAGGGCTTC 712

Qy 271 TCTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCCCGGGGGGCACC 330

Db 711 GCCCTGGAGGCGGCCCTGGCGGTGCGCATGATGTTGCTGGCGGCCTCGCCGCGGCACC 652

Qy 331 ATCTCTAACATTTTACCTTCTGGGTGATGGAGATATGGATCTCAGCATCAGTATGACA 390

Db 651 ACCGCCAACCTCTACAGCCACCTGGCGCATGGCGACGTGGCACTGAACATCACCTTGACC 592

Qy 391 ACCTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATCTCT 442

Db 591 GCGGTGAACTCGGTGATCGCCATCCTCACCATGCCGCTGATCGTCAATCTGT 540

RESULT 11

US-09-833-381-318

; Sequence 318, Application US/09833381
; Patent No. 6672186
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. 6672186el Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 318
; LENGTH: 374
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(374)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-381-318

Query Match 5.2%; Score 58.6; DB 4; Length 374;
Best Local Similarity 57.8%; Pred. No. 1.7e-08;
Matches 141; Conservative 0; Mismatches 100; Indels 3; Gaps 2;

Qy 80 ATGGAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGT 139
||| || || | || | || | || | || | || | |||
Db 131 ATGCAATTCTCAATACAGTGATGAGCACTGTGCTCACCATCCTCTTAGCCATGGTGATGT 190
Qy 140 TCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCGCACATCAGGAGACCCTGGG 199
| ||| |||| || |||| ||| ||| | || || | ||||| ||||
Db 191 TTTCTATGGGTGCAATGTGGAAGTCCACAAGTTCCTAGGACATATAAAGAGACCATGGG 250
Qy 200 GCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCTTT--TACAGCTTATCTCCT 257
| || |||| | |||| ||||| ||||| || || | |||
Db 251 GTATCTTCGTGGGCTTCTCTGTCAAGTTTGAATCATGCCTCTCCACAAGGCTTTTATCC 310
Qy 258 GGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCT-GTTCTCATCATGGGCTGCT 316
| | || | | | || | || | || | || | || | ||||| ||||
Db 311 TGTCTGTGGCCTCTGNATCCTTCTGTACAGGCTGTAGTTGGTGCTAATTATGGGTTGCT 370
Qy 317 GCCC 320
|||
Db 371 GCCC 374

RESULT 12

US-09-540-236-963

; Sequence 963, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
MORAXELLA CATARRHALIS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 963
; LENGTH: 972
; TYPE: DNA
; ORGANISM: M.catarrhalis
US-09-540-236-963

Query Match 5.1%; Score 58.2; DB 4; Length 972;
Best Local Similarity 48.4%; Pred. No. 4.2e-08;
Matches 162; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

Qy	122	TGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCGC	181
Db	152	TGCTTGGCATCGTCATGCTTGGCATGGGTTTGACCTTGACTTTCAAAGATTTTGGTGAAG	211
Qy	182	ACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCCT	241
Db	212	TCACCAAAACCCCAAGGCGGTGATTGTTGGCGTTATCCTTCAATATGTTGTGATGCCAG	271
Qy	242	TTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTTC	301
Db	272	TCATTGCCTTTTTTGTGGTTCAAGCATTAGGCTACCACTGATTTGGCTATCGGTGTCA	331
Qy	302	TCATCATGGGCTGCTGCCCCGGGGGACCATCTCTAACATTTTACCTTCTGGGTTGATG	361
Db	332	TCTTAGTCGGCTGCTGCCCTGGCGGCACCTCGTCAAATGTCATCACTTTTCTTGCCAAAG	391
Qy	362	GAGATATGGATCTCAGCATCAGTATGACAACCTGTTCACCGTGGCCGCCCTGGGAATGA	421
Db	392	GCAATACCGCTTTATCAGTTGCTTGCACGACACTCTCCACACTCTTAGCCCCTATTTTGA	451
Qy	422	TGCCACTCTGCATTTATCTCTACACCTGGTCCTGG	456
Db	452	CACCAGCTGTATTTTATTTATTTGCCAGCCAATGG	486

RESULT 13

US-09-596-002-41/c

; Sequence 41, Application US/09596002

; Patent No. 6632636

; GENERAL INFORMATION:

; APPLICANT: Lagace, Robert, E.

; APPLICANT: Patterson, Chandra

; APPLICANT: Berg, Kim, L.

; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME

; FILE REFERENCE: PM-0008-4 US

; CURRENT APPLICATION NUMBER: US/09/596,002

; CURRENT FILING DATE: 2000-06-16

; PRIOR APPLICATION NUMBER: 60/140,121

; PRIOR FILING DATE: 1999-06-18

; NUMBER OF SEQ ID NOS: 41

US-09-596-002-41

Matches 162; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

183179

183119

183059

182999

182939

Db 182938 CGCCAGCTGTATTTTATTTATTTGCCAGCCAATGG 182904

RESULT 14

US-09-252-991A-1376/c

; Sequence 1376, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1376
; LENGTH: 891
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1376

Query Match 4.9%; Score 55.8; DB 4; Length 891;
Best Local Similarity 47.1%; Pred. No. 2.3e-07;
Matches 171; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

```
Qy      94 CTCGTTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGTTCTTTTGGGATGT 153
      ||| |||| | | | | | | | | | | | | | | | |
Db      620 CTCCCGCTCACCGCAGCCATCGCGCCACTGCTCGGCCTGGTGTGATGTTTCGGCATGGGCTG 561

Qy      154 TCCGTGGAGATCCGGAAGCTGTGGTTCGCACATCAGGAGACCCTGGGGCATTGCTGTGGGA 213
      | | | | | | | | | | | | | | | | | | | |
Db      560 ACGCTCAAGGGCGAAGACTTCCGCGAGGTCGCGCCGACCCCATACGGGTGCTGATCGGC 501

Qy      214 CTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGGCCATTAGCTTTTCT 273
      |||| ||||| | ||||| | | | | | | | |
Db      500 GTGCTGGCCAGTTCGTCATCATGCCCCGGCCTGGCCTGGTGTGCTCTGCAGCCTGTTGCAG 441

Qy      274 CTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCCCCGGGGGGCACCATC 333
      || | | | | | | | | | | | | | | | | | |
Db      440 TTGCCGGCGGAGATCGCGGTGGGCGTGATCCTGGTCGGCTGCTGCCCCGGCGGCACCGCT 381

Qy      334 TCTAACATTTTCACCTTCTGGGTTGATGGAGATATGGATCTCAGCATCAGTATGACAACC 393
      || ||| | | |||| | | | | | | | | | | | |
Db      380 TCCAACGTGATGACCTGGCTGTCCCGTGGCGATGTCGCCCTGTCGGTGGCGATCACCTCG 321

Qy      394 TGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATCTCTACACCTGGTCC 453
      ||||| || ||||| | | | | | | | | | | | |
Db      320 GTGACCACCTGCTCGCCCCGCTGGTACGCCGGCGCTGGTCTGGCTGCTGGCTTCGGCC 261

Qy      454 TGG 456
      |||
Db      260 TGG 258
```

RESULT 15

US-09-252-991A-1144

; Sequence 1144, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 1144
 ; LENGTH: 948
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-1144

Query Match 4.9%; Score 55.8; DB 4; Length 948;
 Best Local Similarity 47.1%; Pred. No. 2.4e-07;
 Matches 171; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

```

Qy      94 CTCGTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGTTCTCTTTGGGATGT 153
      ||| |||| | | | | | | | | | | | | | | | |
Db     112 CTCCCGCTACCGCAGCCATCGCGCCACTGCTCGGCCCTGGTGATGTTTCGGCATGGGCCTG 171

Qy     154 TCCGTGGAGATCCGGAAGCTGTGGTTCGCACATCAGGAGACCCTGGGGCATTGCTGTGGGA 213
      | | | | | | | | | | | | | | | | | |
Db     172 ACGCTCAAGGGCGAAGACTTCCGCGAGGTGCGCCGGCACCCCATACGGGTGCTGATCGGC 231

Qy     214 CTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGGCCATTAGCTTTTCT 273
      |||| ||||| | ||||| | | | | | | |
Db     232 GTGCTGGCCAGTTCGTCATCATGCCCGGCCTGGCCTGGTTGCTCTGCAGCCTGTTGCAG 291

Qy     274 CTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCCCCGGGGGGCACCATC 333
      || | | | | | | | | | | | | | | | |
Db     292 TTGCCGGCGGAGATCGCGGTGGGCGTGATCCTGGTCGGCTGCTGCCCCGGCGGCACCGCT 351

Qy     334 TCTAACATTTTACCTTCTCTGGGTTGATGGAGATATGGATCTCAGCATCAGTATGACAACC 393
      || ||| | | |||| | ||| ||| | | | | | |
Db     352 TCCAACGTGATGACCTGGCTGTCCCGTGGCGATGTCGCCCTGTCGGTGGCGATCACCTCG 411

Qy     394 TGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATCTCTACACCTGGTCC 453
      ||||| || ||||| | | ||| | | | | | |
Db     412 GTGACCACCCTGCTCGCCCCGCTGGTCACGCCGGCGCTGGTCTGGCTGCTGGCTTCGGCC 471

Qy     454 TGG 456
      |||
Db     472 TGG 474
  
```

Search completed: March 25, 2004, 18:55:43
 Job time : 131 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2004, 16:28:48 ; Search time 491 Seconds
(without alignments)
8598.560 Million cell updates/sec

Title: US-10-091-628-1
Perfect score: 1134
Sequence: 1 atgagagccaattgttccag.....acatcacttcatgtgaatag 1134

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2458946 seqs, 1861504846 residues

Total number of hits satisfying chosen parameters: 4917892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	
No.	Score Match Length DB ID	Description

	1	1134	100.0	1134	13	US-10-091-628-1	Sequence 1, Appli
	2	1134	100.0	1600	13	US-10-091-628-3	Sequence 3, Appli
	3	655.8	57.8	987	11	US-09-981-151A-11	Sequence 11, Appl
	4	297.8	26.3	3779	15	US-10-191-997-110	Sequence 110, App
	5	183.2	16.2	1663	9	US-09-917-800A-1626	Sequence 1626, Ap
	6	183.2	16.2	1663	15	US-10-388-934-263	Sequence 263, App
	7	173.6	15.3	1580	9	US-09-880-107-2176	Sequence 2176, Ap
	8	173.6	15.3	1580	14	US-10-288-222A-15	Sequence 15, Appl
	9	141.4	12.5	1988	15	US-10-085-198-113	Sequence 113, App
c	10	79.8	7.0	360	9	US-09-864-761-31375	Sequence 31375, A
c	11	79.8	7.0	560	9	US-09-864-761-14847	Sequence 14847, A
c	12	77.8	6.9	310	9	US-09-833-381-317	Sequence 317, App
	13	74	6.5	401	9	US-09-960-352-2253	Sequence 2253, Ap
	14	67.6	6.0	972	9	US-09-738-626-2554	Sequence 2554, Ap
c	15	67.6	6.0	3309400	9	US-09-738-626-1	Sequence 1, Appli
	16	64.4	5.7	1017	9	US-09-938-842A-380	Sequence 380, App
	17	64.4	5.7	1017	11	US-09-938-842A-380	Sequence 380, App
	18	62.4	5.5	1152	15	US-10-093-463-21	Sequence 21, Appl
	19	62.4	5.5	1152	15	US-10-093-463-25	Sequence 25, Appl
	20	62.4	5.5	1317	13	US-10-091-628-4	Sequence 4, Appli
	21	62.4	5.5	1355	15	US-10-093-463-23	Sequence 23, Appl
	22	62.4	5.5	1777	13	US-10-091-628-6	Sequence 6, Appli
	23	62.4	5.5	2520	15	US-10-108-260A-919	Sequence 919, App
	24	58.6	5.2	374	9	US-09-833-381-318	Sequence 318, App
	25	54.2	4.8	407	9	US-09-960-352-10081	Sequence 10081, A
	26	53.6	4.7	1005	9	US-09-738-626-1392	Sequence 1392, Ap
	27	53.6	4.7	3309400	9	US-09-738-626-1	Sequence 1, Appli
	28	53.2	4.7	418	9	US-09-960-352-4473	Sequence 4473, Ap
c	29	44.6	3.9	197	9	US-09-864-761-30128	Sequence 30128, A
c	30	44.6	3.9	600	9	US-09-864-761-13589	Sequence 13589, A
	31	43.2	3.8	1425	10	US-09-796-753-61	Sequence 61, Appl
	32	42.6	3.8	912	9	US-09-974-300-6823	Sequence 6823, Ap
	33	42.4	3.7	479	12	US-10-424-599-41790	Sequence 41790, A
c	34	39	3.4	592	10	US-09-902-563-5	Sequence 5, Appli
c	35	39	3.4	592	14	US-10-096-255-5	Sequence 5, Appli
c	36	39	3.4	5403	10	US-09-902-563-3	Sequence 3, Appli
c	37	39	3.4	5403	14	US-10-096-255-3	Sequence 3, Appli
	38	37.8	3.3	65	10	US-09-908-975-26629	Sequence 26629, A
	39	37.8	3.3	1845	12	US-10-282-122A-31931	Sequence 31931, A
	40	37.4	3.3	738	12	US-10-424-599-34013	Sequence 34013, A
c	41	36.8	3.2	3295	12	US-10-383-241B-5	Sequence 5, Appli
	42	36.6	3.2	1306	12	US-10-424-599-79635	Sequence 79635, A
	43	36.4	3.2	1275	14	US-10-113-113-3	Sequence 3, Appli
c	44	36	3.2	2457	15	US-10-094-749-438	Sequence 438, App
c	45	36	3.2	2524	10	US-09-814-353-21076	Sequence 21076, A

ALIGNMENTS

RESULT 1

US-10-091-628-1

; Sequence 1, Application US/10091628

; Publication No. US20020164627A1

; GENERAL INFORMATION:

; APPLICANT: Wilganowski, Nathaniel L.

```
; APPLICANT: Nepomnichy, Boris
; APPLICANT: Burnett, Michael B.
; APPLICANT: Hu, Yi
; TITLE OF INVENTION: No. US20020164627A1el Human Transporter Proteins and
Polynucleotides Encoding the
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0314-USA
; CURRENT APPLICATION NUMBER: US/10/091,628
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: US 60/275,009
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/284,152
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1134
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-091-628-1
```

```
Query Match          100.0%; Score 1134; DB 13; Length 1134;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG 60
        |||
Db      1 ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG 60

Qy     61 CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTG 120
        |||
Db     61 CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTG 120

Qy    121 ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTTCG 180
        |||
Db    121 ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTTCG 180

Qy    181 CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT 240
        |||
Db    181 CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT 240

Qy    241 TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT 300
        |||
Db    241 TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT 300

Qy    301 CTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGAT 360
        |||
Db    301 CTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGAT 360

Qy    361 GGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATG 420
        |||
Db    361 GGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATG 420

Qy    421 ATGCCACTCTGCATTTATCTCTACACCTGGTCTGGAGTCTTCAGCAGAATCTCACCATT 480
        |||
Db    421 ATGCCACTCTGCATTTATCTCTACACCTGGTCTGGAGTCTTCAGCAGAATCTCACCATT 480
```

Qy 481 CCTTATCAGAACATAGGAATTACCCTTGTGTGCCTGACCATTCTGTGGCCTTTGGTGTC 540
 |||
 Db 481 CCTTATCAGAACATAGGAATTACCCTTGTGTGCCTGACCATTCTGTGGCCTTTGGTGTC 540

Qy 541 TATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTT 600
 |||
 Db 541 TATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTT 600

Qy 601 GGTGGGGTCTCCTTCTGGTGGTTCGAGTTGCTGGTGTGGTCTGGCGAAAGGATCTTGG 660
 |||
 Db 601 GGTGGGGTCTCCTTCTGGTGGTTCGAGTTGCTGGTGTGGTCTGGCGAAAGGATCTTGG 660

Qy 661 AATTCAGACATCACCCCTTCTGACCATCAGTTTCATCTTTCCTTTGATTGGCCATGTCACG 720
 |||
 Db 661 AATTCAGACATCACCCCTTCTGACCATCAGTTTCATCTTTCCTTTGATTGGCCATGTCACG 720

Qy 721 GGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTCCTTA 780
 |||
 Db 721 GGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTCCTTA 780

Qy 781 GAAACTGGAGCTCAGAATATTCAGATGTGCATCACCATGCTCCAGTTATCTTTCAGTCT 840
 |||
 Db 781 GAAACTGGAGCTCAGAATATTCAGATGTGCATCACCATGCTCCAGTTATCTTTCAGTCT 840

Qy 841 GAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGAT 900
 |||
 Db 841 GAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGAT 900

Qy 901 GGATTTCTTATTGTTGCAGCATATCAGACGTACAAGAGGAGATTGAAGAACAACATGGA 960
 |||
 Db 901 GGATTTCTTATTGTTGCAGCATATCAGACGTACAAGAGGAGATTGAAGAACAACATGGA 960

Qy 961 AAAAAGAACTCAGGTTGCACAGAAGTCTGCCATACGAGGAAATCGACTTCTTCCAGAGAG 1020
 |||
 Db 961 AAAAAGAACTCAGGTTGCACAGAAGTCTGCCATACGAGGAAATCGACTTCTTCCAGAGAG 1020

Qy 1021 ACCAATGCCTTCTTGGAGGTGAATGAAGAAGGTGCCATCACTCCTGGGCCACCAGGGCCA 1080
 |||
 Db 1021 ACCAATGCCTTCTTGGAGGTGAATGAAGAAGGTGCCATCACTCCTGGGCCACCAGGGCCA 1080

Qy 1081 ATGGATTGCCACAGGGCTCTCGAGCCAGTTGGCCACATCACTTCATGTGAATAG 1134
 |||
 Db 1081 ATGGATTGCCACAGGGCTCTCGAGCCAGTTGGCCACATCACTTCATGTGAATAG 1134

RESULT 2

US-10-091-628-3

; Sequence 3, Application US/10091628

; Publication No. US20020164627A1

; GENERAL INFORMATION:

; APPLICANT: Wilganowski, Nathaniel L.

; APPLICANT: Nepomnichy, Boris

; APPLICANT: Burnett, Michael B.

; APPLICANT: Hu, Yi

; TITLE OF INVENTION: No. US20020164627A1el Human Transporter Proteins and Polynucleotides Encoding the

; TITLE OF INVENTION: Same

Db 734 TATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATTGGGGCCGTTGTT 793

Qy 601 GGTGGGGTCTCCTTCTGGTGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGATCTTGG 660
 |||

Db 794 GGTGGGGTCTCCTTCTGGTGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGATCTTGG 853

Qy 661 AATTCAGACATCACCTTCTGACCATCAGTTTCATCTTTCCTTTGATTGGCCATGTCACG 720
 |||

Db 854 AATTCAGACATCACCTTCTGACCATCAGTTTCATCTTTCCTTTGATTGGCCATGTCACG 913

Qy 721 GGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCTTA 780
 |||

Db 914 GGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCTTA 973

Qy 781 GAAACTGGAGCTCAGAATATTCAGATGTGCATCACCATGCTCCAGTTATCTTTCACTGCT 840
 |||

Db 974 GAAACTGGAGCTCAGAATATTCAGATGTGCATCACCATGCTCCAGTTATCTTTCACTGCT 1033

Qy 841 GAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGAT 900
 |||

Db 1034 GAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGAT 1093

Qy 901 GGATTTCTTATTGTTGCAGCATATCAGACGTACAAGAGGAGATTGAAGAACAAACATGGA 960
 |||

Db 1094 GGATTTCTTATTGTTGCAGCATATCAGACGTACAAGAGGAGATTGAAGAACAAACATGGA 1153

Qy 961 AAAAAGAACTCAGGTTGCACAGAAGTCTGCCATACGAGGAAATCGACTTCTTCCAGAGAG 1020
 |||

Db 1154 AAAAAGAACTCAGGTTGCACAGAAGTCTGCCATACGAGGAAATCGACTTCTTCCAGAGAG 1213

Qy 1021 ACCAATGCCTTCTTGGAGGTGAATGAAGAAGGTGCCATCACTCCTGGGCCACCAGGGCCA 1080
 |||

Db 1214 ACCAATGCCTTCTTGGAGGTGAATGAAGAAGGTGCCATCACTCCTGGGCCACCAGGGCCA 1273

Qy 1081 ATGGATTGCCACAGGGCTCTCGAGCCAGTTGGCCACATCACTTCATGTGAATAG 1134
 |||

Db 1274 ATGGATTGCCACAGGGCTCTCGAGCCAGTTGGCCACATCACTTCATGTGAATAG 1327

RESULT 3

US-09-981-151A-11

; Sequence 11, Application US/09981151A

; Publication No. US20030212256A1

; GENERAL INFORMATION:

; APPLICANT: Edinger, Shlomit R
 ; APPLICANT: Gerlach, Valerie
 ; APPLICANT: MacDougall, John R
 ; APPLICANT: Malyankar, Muriel M
 ; APPLICANT: Smithson, Glenda
 ; APPLICANT: Millet, Isabelle
 ; APPLICANT: Peyman, John A
 ; APPLICANT: Stone, David J
 ; APPLICANT: Gunther, Erik
 ; APPLICANT: Ellerman, Karen
 ; APPLICANT: Shimkets, Richard A
 ; APPLICANT: Padigaru, Muralidhara
 ; APPLICANT: Guo, Xiaojia

```

; APPLICANT: Patturajan, Meera
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Burgess, Catherine E
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Gangolli, Esha A
; APPLICANT: Fernandes, Elma R
; APPLICANT: Gorman, Linda
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-168
; CURRENT APPLICATION NUMBER: US/09/981,151A
; CURRENT FILING DATE: 2001-10-16
; PRIOR APPLICATION NUMBER: 60/241,040
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/241,058
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/241,063
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/241,243
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/242,152
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/242,482
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/242,611
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/242,612
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/242,880
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: 60/242,881
; PRIOR FILING DATE: 2000-10-24
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 987
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-981-151A-11

```

```

Query Match          57.8%;  Score 655.8;  DB 11;  Length 987;
Best Local Similarity 86.3%;  Pred. No. 4.5e-205;
Matches 803;  Conservative 0;  Mismatches 77;  Indels 51;  Gaps 5;

```

```

Qy      1 ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGAGAGCCAATTGTTCCAGCAGCTCAGCCTGCCCTGCCAACAGTTCAGAGGAGGAGCTG 60

Qy     61 CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTG 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCAGTGGGACTGGAGGTGCATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTATC 120

Qy    121 ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCG 180
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCG 180

```

Qy 181 CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT 240
 |||
 Db 181 CACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCT 240

Qy 241 TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT 300
 |||
 Db 241 TTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTT 300

Qy 301 CTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTGAT 360
 |||
 Db 301 CTCATCATGGGCTGCTG-CCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTGAT 359

Qy 361 GGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATG 420
 |||
 Db 360 GGAGATATGGATCTCA-----GGTGCCCTGGGAATG 390

Qy 421 ATGCCACTCTGCATTTATCTCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATT 480
 |||
 Db 391 ATGCCACTCTGCATTTATCTCTACACCTGGTCCTGGAGTCTTCAGCAGAATCTCACCATT 450

Qy 481 CCTTATCAGAACA-----TAGGAATTACCCTTGTGTGCCTGACCATTCCCTGTG 528
 |||
 Db 451 CCTTATCAGAACATAGGTCTGTCTTTAGGAATTACCCTTGTGTGCCTGACCATTCCCTGTG 510

Qy 529 GCCTTTGGTGTCTATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATT 588
 |||
 Db 511 GCCTTTGGTGTCTATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAA---- 566

Qy 589 GGGGCCGTTGTTGGTGGGGTCCCTCTCTGGTGGTGCAGTTGCTGGTGTGGTCCCTGGCG 648
 |||
 Db 567 --GGCCGTTGTTGGTGGGGTCCCTCTCTGGTGGTGCAGTTGCTGGTGTGGTCCCTGGCG 624

Qy 649 AAAGGATCTTGGAATTCAGACATCACCTTCTGACCATCAGTTTCATCTTTCCTTTGATT 708
 |||
 Db 625 AAAGGATCTTGGAATTCAGACATCACCTTCTGACCATCAGTTTCATCTTTCCTTTGATT 684

Qy 709 GGCCATGTCACGGGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGG 768
 |||
 Db 685 GGCCATGTCACGGGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGACCTTG 744

Qy 769 ACAATTTCTTTAGAACTGGAGCTCAGAATATTAGATGTGCATCACCATGCTCCAGTTA 828
 |||
 Db 745 CCTATCTTTTAG---GTTTAGCTTTCAAGACACCCTGTGATACCCTACTCGCAATGACT 801

Qy 829 TCTTTCACTGCTGAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTC 888
 |||
 Db 802 TCGTGTCTGAATGTTCCAGGCTCATCTATGCCTTCATTCCCTCTGCTATATGGACTCTTC 861

Qy 889 CAGCTGATAGATGGATTTCTTATTGTTGCAG 919
 |||
 Db 862 CAGCTGATAGATGGATTTCTTATTGTTGAAG 892

RESULT 4

US-10-191-997-110

; Sequence 110, Application US/10191997

```
; Publication No. US20030207834A1
; GENERAL INFORMATION:
; APPLICANT: Oligos Etc., Inc.
; APPLICANT: DALE, Roderic M. K.
; APPLICANT: ARROW, Amy
; APPLICANT: THOMPSON, Terry
; TITLE OF INVENTION: Oligonucleotide-Containing Pharmacological Compositions
And Their Use
; FILE REFERENCE: 54800-5019
; CURRENT APPLICATION NUMBER: US/10/191,997
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: US 60/303,820
; PRIOR FILING DATE: 2001-07-10
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 110
; LENGTH: 3779
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: IBAT: Acc. No. US20030207834A1 NM_000452
US-10-191-997-110
```

```
Query Match          26.3%; Score 297.8; DB 15; Length 3779;
Best Local Similarity 58.5%; Pred. No. 1.3e-86;
Matches 518; Conservative 0; Mismatches 367; Indels 0; Gaps 0;
```

```
Qy      80 ATGGAAACCTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGT 139
      || | ||| | || | | ||| || | || ||| || | |||
Db      678 ATAACATCCTAAGTGTGGTCCTAAGTACGGTGTGACCATCCTGTTGGCCTTGGTGATGT 737

Qy      140 TCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCGCACATCAGGAGACCCTGGG 199
      |||| ||||| ||||| ||| ||| | ||||| | | | |||
Db      738 TCTCCATGGGATGCAACGTGGAAATCAAGAAATTTCTAGGGCACATAAAGCGGCCCGTGGG 797

Qy      200 GCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGG 259
      ||||| ||| || | |||| ||||| ||||| | |||| | ||||
Db      798 GCATTTGTGTTGGCTTCCTCTGTCAAGTTTGAATCATGCCCTCACAGGATTCATCCTGT 857

Qy      260 CCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCATGGGCTGCTGCC 319
      | | |||| | || |||| || | | || |||| || || |||||
Db      858 CGGTGGCCTTTGACATCCTCCCGCTCCAGGCCGTAGTGGTGTCTATTATAGGATGCTGCC 917

Qy      320 CGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGATGGAGATATGGATCTCAGCA 379
      | || || || || || || || || || || || || || || ||
Db      918 CTGGAGGAACTGCCTCCAATATCTTGGCCTATTGGGTCGATGGCGACATGGACCTGAGCG 977

Qy      380 TCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACTCTGCATTTATC 439
      |||| |||| || || |||| || || |||| ||||| || ||| |
Db      978 TCAGCATGACCACATGCTCCACACTGCTTGGCCTCGGAATGATGCCGCTGTGCCTCCTTA 1037

Qy      440 TCTACACCTGGTCTCTGGAGTCTTCAGCAGAATCTCACCATTCTTATCAGAACATAGGAA 499
      |||| ||| ||| || || || || || || || || || || ||
Db      1038 TCTATACCAAAATGTGGGTGCGACTCTGGGAGCATCGTAATTCCTATGATAACATAGGTA 1097

Qy      500 TTACCCTTGTGTGCCTGACCATTCCTGTGGCCTTTGGTGTCTATGTGAATTACAGATGGC 559
```


Db	1098	CATCTCTGGTTGCTCTCGTTGTTCCATTGGAATGTTTGTTAATCACAAATGGC	1157
Qy	560	CAAAACAATCCAAAATCATTTCTCAAGATTGGGGCCGTTGTTGGTGGGGTCCTCCTTCTGG	619
Db	1158	CCCAAAAAGCAAAGATCATACTTAAATTTGGGTCCATCGCGGGCGCCATCCTCATTGTGC	1217
Qy	620	TGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGATCTTGGAATTCAGACATCACCCCTTC	679
Db	1218	TCATAGCTGTGGTTGGAGGAATATTGTACCAAAGCGCCTGGATCATTGCTCCCAAACGT	1277
Qy	680	TGACCATCAGTTTTCATCTTTCCTTTGATTGGCCATGTACGGGTTTTCTGCTGGCACTTT	739
Db	1278	GGATTATAGGAACAATATTTCTGTGGCGGGTACTCCCTGGGGTTTTCTTCTGGCTAGAA	1337
Qy	740	TTACCCACCAGTCTTGGCAAAGGTGCAGGACAATTTCTTAGAAACTGGAGCTCAGAATA	799
Db	1338	TTGCTGGTCTACCCTGGTACAGGTGCCGAACGGTTGCTTTTGAAACGGGGATGCAGAACA	1397
Qy	800	TTCAGATGTGCATCACCATGCTCCAGTTATCTTTCACTGCTGAGCACTTGGTCCAGATGT	859
Db	1398	CGCAGCTATGTTCCACCATCGTTCAGCTCTCCTTCACTCCTGAGGAGCTCAATGTCGTAT	1457
Qy	860	TGAGTTTCCCACTGGCCTATGGACTCTTCCAGCTGATAGATGGATTTCTTATTGTTGCAG	919
Db	1458	TCACCTTCCCGCTCATCTACAGCATTTTCCAGCTCGCCTTTGCCGCAATATTCTTAGGAT	1517
Qy	920	CATATCAGACGTACAAGAGGAGATTGAAGAACAAACATGGAAAAA	964
Db	1518	TTTATGTGGCATAACAAGAAATGTCATGGAAAAACAAGGCAGAAA	1562

RESULT 5

US-09-917-800A-1626

; Sequence 1626, Application US/09917800A

; Patent No. US20020119462A1

; GENERAL INFORMATION:

; APPLICANT: Mendrick, Donna

; APPLICANT: Porter, Mark

; APPLICANT: Johnson, Kory

; APPLICANT: Castle, Arthur

; APPLICANT: Elashoff, Michael

; APPLICANT: Gene Logic, Inc.

; TITLE OF INVENTION: Molecular Toxicology Modeling

; FILE REFERENCE: 44921-5038-US

; CURRENT APPLICATION NUMBER: US/09/917,800A

; CURRENT FILING DATE: 2001-07-31

; PRIOR APPLICATION NUMBER: US 60/222,040

; PRIOR FILING DATE: 2000-07-31

; PRIOR APPLICATION NUMBER: US 60/222,880

; PRIOR FILING DATE: 2000-11-02

; PRIOR APPLICATION NUMBER: US 60/290,029

; PRIOR FILING DATE: 2001-05-11

; PRIOR APPLICATION NUMBER: US 60/290,645

; PRIOR FILING DATE: 2001-05-15

; PRIOR APPLICATION NUMBER: US 60/292,336

; PRIOR FILING DATE: 2001-05-22

```
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1626
; LENGTH: 1663
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM_017047
US-09-917-800A-1626
```

```
Query Match          16.2%; Score 183.2; DB 9; Length 1663;
Best Local Similarity 53.6%; Pred. No. 5.1e-49;
Matches 430; Conservative 0; Mismatches 363; Indels 9; Gaps 2;
```

```
Qy      119 TGATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGT 178
      | | | | | | | | | | | | | | | | | | | | | | | |
Db      219 TAATGTTGCTGCTTATCATGCTCTCACTGGGCTGCACCATGGAATTCAGCAAGATCAAGG 278

Qy      179 CGCACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGC 238
      | | | | | | | | | | | | | | | | | | | | | | | |
Db      279 CTCACCTGTGGAAGCCCAAAGGGGTGATCGTTGCCTTGGTGGCCCAGTTTGGCATCATGC 338

Qy      239 CTTTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTG 298
      | | | | | | | | | | | | | | | | | | | | | | | |
Db      339 CCCTCGCTGCTTTTCTTCTCGGCAAGATCTTTCACCTGAGCAACATTGAAGCTCTGGCCA 398

Qy      299 TTCTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTG 358
      | | | | | | | | | | | | | | | | | | | | | | | |
Db      399 TCCTCATCTGTGGCTGCTCTCCCGGGGGGAACTTGTCCAACCTCTTCACCCTGGCCATGA 458

Qy      359 ATGGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAA 418
      | | | | | | | | | | | | | | | | | | | | | | | |
Db      459 AGGGGGACATGAACCTCAGCATCGTGATGACCACCTGCTCCAGCTTCAGTGCCTTGGGCA 518

Qy      419 TGATGCCACTCTGCATTTATCTCTACACC---TGGTCCTGGAGTCTTCAGCAGAATCTCA 475
      | | | | | | | | | | | | | | | | | | | | | | | |
Db      519 TGATGCCACTCCTCTTATACGTCTACAGCAAAGGCATCTACGATGGAGACCTTAAGGACA 578

Qy      476 CCATTTCCTTATCAGAACATAGGAATTACCCTTGTGTGCCTGACCATTCTGTGGCCTTTG 535
      | | | | | | | | | | | | | | | | | | | | | | | |
Db      579 AGGTGCCCTACAAAGGCATTATGATATCACTAGTCATAGTTCTCATTCCTTGCACCATAG 638

Qy      536 GTGTCTATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATTGGGGCCG 595
      | | | | | | | | | | | | | | | | | | | | | | | |
Db      639 GGATCGTCCTCAAGTCCAAAAGGCCACACTATGTACCCTACATCCTCAAGGGAGGCATGA 698

Qy      596 TTGTTGGTGGGGTCCTCCTTCTGGTGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGAT 655
      | | | | | | | | | | | | | | | | | | | | | | | |
Db      699 TCATCACCTTCCTCCTCTCTGTGGCTGTACAGCCCTCTCTGTATCAATGTGGGCAACA 758
```

Qy 656 CTTGGAATTCAGACATCAC-----CCTTCTGACCATCAGTTTCATCTTTCCTTTGATTG 709
 | | | | | | | | | | | | | | | | | |
 Db 759 GCATCATGTTTCGTCATGACACCACACTTACTGGCTACCTCCTCCCTGATGCCCTTCTCTG 818
 Qy 710 GCCATGTCACGGGTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGGA 769
 | | | | | | | | | | | | | | | | | |
 Db 819 GCTTTCTGATGGGTTACATTCTCTCTGCTCTCTTCCAACCTCAATCCAAGCTGCAGACGCA 878
 Qy 770 CAATTTCTTAGAACTGGAGCTCAGAATATTCAGATGTGCATCACCATGCTCCAGTTAT 829
 | | | | | | | | | | | | | | | | | |
 Db 879 CCATCAGCATGGAAACAGGATTCCAAAACATTCAACTCTGTTCTACCATCCTCAATGTGA 938
 Qy 830 CTTTCACTGCTGAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTCC 889
 | | | | | | | | | | | | | | | | | |
 Db 939 CCTTCCCCCTGAAGTCATTGGGCCACTTTTCTTCTTTCCTCTCTCTACATGATTTTCC 998
 Qy 890 AGCTGATAGATGGATTTCTTAT 911
 | | | | | | | | | | | | | | | | | |
 Db 999 AGCTTGCAGAAGGACTTCTCAT 1020

RESULT 6

US-10-388-934-263

; Sequence 263, Application US/10388934

; Publication No. US20040005547A1

; GENERAL INFORMATION:

; APPLICANT: Boess, Franziska

; APPLICANT: Suter-Dick, Laura

; APPLICANT: Wolf, Detlef

; TITLE OF INVENTION: BIOMARKERS AND EXPRESSION PROFILES FOR TOXICOLOGY

; FILE REFERENCE: 21199

; CURRENT APPLICATION NUMBER: US/10/388,934

; CURRENT FILING DATE: 2003-03-14

; PRIOR APPLICATION NUMBER: 02005336.9

; PRIOR FILING DATE: 2002-03-14

; PRIOR APPLICATION NUMBER: 02015657.6

; PRIOR FILING DATE: 2002-07-17

; NUMBER OF SEQ ID NOS: 862

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 263

; LENGTH: 1663

; TYPE: DNA

; ORGANISM: Rattus norvegicus (No. US20040005547A1way rat)

US-10-388-934-263

Query Match 16.2%; Score 183.2; DB 15; Length 1663;
 Best Local Similarity 53.6%; Pred. No. 5.1e-49;
 Matches 430; Conservative 0; Mismatches 363; Indels 9; Gaps 2;

Qy 119 TGATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGT 178
 | | | | | | | | | | | | | | | | | |
 Db 219 TAATGTTGCTGCTTATCATGCTCTCACTGGGCTGCACCATGGAATTCAGCAAGATCAAGG 278
 Qy 179 CGCACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGC 238
 | | | | | | | | | | | | | | | | | |
 Db 279 CTCACTTGTGGAAGCCCAAAGGGGTGATCGTTGCCTTGGTGGCCCAGTTTGGCATCATGC 338

Qy 239 CTTTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTG 298
 | | | | | | | | | | | | | | | | | |
 Db 339 CCCTCGCTGCTTTTCTTCTCGGCAAGATCTTTACCTGAGCAACATTGAAGCTCTGGCCA 398

Qy 299 TTCTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTG 358
 | | | | | | | | | | | | | | | | | |
 Db 399 TCCTCATCTGTGGCTGCTCTCCCGGGGGGAACCTGTCCAACCTCTTCACCCTGGCCATGA 458

Qy 359 ATGGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAA 418
 | | | | | | | | | | | | | | | | | |
 Db 459 AGGGGGACATGAACCTCAGCATCGTGATGACCACCTGCTCCAGCTTCAGTGCCTTGGGCA 518

Qy 419 TGATGCCACTCTGCATTTATCTCTACACC---TGGTCCTGGAGTCTTCAGCAGAATCTCA 475
 | | | | | | | | | | | | | | | | | |
 Db 519 TGATGCCACTCCTCTTATACGTCTACAGCAAAGGCATCTACGATGGAGACCTTAAGGACA 578

Qy 476 CCATTCCCTTATCAGAACATAGGAATTACCCTTGTGTGCCTGACCATTCTGTGGCCTTTG 535
 | | | | | | | | | | | | | | | | | |
 Db 579 AGGTGCCCTACAAAGGCATTATGATATCACTAGTCATAGTTCTCATTCTTGCACCATAG 638

Qy 536 GTGTCTATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATTGGGGCCG 595
 | | | | | | | | | | | | | | | | | |
 Db 639 GGATCGTCCTCAAGTCCAAAAGGCCACACTATGTACCCTACATCCTCAAGGGAGGCATGA 698

Qy 596 TTGTTGGTGGGGTCCTCCTTCTGGTGGTCGCAGTTGCTGGTGTGGTCCTGGCGAAAGGAT 655
 | | | | | | | | | | | | | | | | | |
 Db 699 TCATCACCTTCCTCCTCTCTGTGGCTGTACAGCCCTCTCTGTCATCAATGTGGGCAACA 758

Qy 656 CTTGGAATTCAGACATCAC-----CCTTCTGACCATCAGTTTCATCTTTCCTTTGATTG 709
 | | | | | | | | | | | | | | | | | |
 Db 759 GCATCATGTTTCGTATGACACCACACTTACTGGCTACCTCCTCCCTGATGCCCTTCTCTG 818

Qy 710 GCCATGTCACGGGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGGA 769
 | | | | | | | | | | | | | | | | | |
 Db 819 GCTTTCTGATGGGTTACATTCTCTCTGCTCTTCCAACCTCAATCCAAGCTGCAGACGCA 878

Qy 770 CAATTTTCCTTAGAACTGGAGCTCAGAATATTCAGATGTGCATCACCATGCTCCAGTTAT 829
 | | | | | | | | | | | | | | | | | |
 Db 879 CCATCAGCATGGAAACAGGATTCCAAAACATCAACTCTGTTCTACCATCCTCAATGTGA 938

Qy 830 CTTTCACTGCTGAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTCC 889
 | | | | | | | | | | | | | | | | | |
 Db 939 CCTTCCCCCCTGAAGTCATTGGGCCACTTTTCTTCTTCTCCTCTACATGATTTTCC 998

Qy 890 AGCTGATAGATGGATTTCTTAT 911
 | | | | | | | | | | | |
 Db 999 AGCTTGCAGAAGGACTTCTCAT 1020

RESULT 7

US-09-880-107-2176

; Sequence 2176, Application US/09880107

; Patent No. US20020142981A1

; GENERAL INFORMATION:

; APPLICANT: Horne, Darci T.

; APPLICANT: Vockley, Joseph G.

```
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2176
; LENGTH: 1580
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 L21893
US-09-880-107-2176
```

```
Query Match          15.3%; Score 173.6; DB 9; Length 1580;
Best Local Similarity 51.9%; Pred. No. 7.2e-46;
Matches 445; Conservative 0; Mismatches 404; Indels 9; Gaps 2;
```

```
Qy      119 TGATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGT 178
          | ||| || | |||| |||| |||| || || |||| || | ||| | |
Db      180 TCATGTTGTTCTTCATCATGCTCTCGCTGGGCTGCACCATGGAGTTCAGCAAGATCAAGG 239

Qy      179 CGCACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGGCTCATGC 238
          | ||| | ||| || || || || || || || || || || || || || || || ||
Db      240 CTCACTTATGGAAGCCTAAAGGGCTGGCCATCGCCCTGGTGGCACAGTATGGCATCATGC 299

Qy      239 CTTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTG 298
          | | || || | | |||| || || || || || || || || || || || || ||
Db      300 CCCTCACGGCCTTTGTGCTGGGCAAGGTCTTCCGGCTGAAGAACATTGAGGCACTGGCCA 359

Qy      299 TTCTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTG 358
          | | || || |||| || || || || || || || || || || || || || || ||
Db      360 TCTTGGTCTGTGGCTGCTCACCTGGAGGGAACCTGTCCAATGTCTTCAGTCTGGCCATGA 419

Qy      359 ATGGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAA 418
          | || || || | |||| || || || || || || || || || || || || || ||
Db      420 AGGGGGACATGAACCTCAGCATTGTGATGACCACCTGCTCCACCTTCTGTGCCCTTGGCA 479

Qy      419 TGATGCCACTCTGCATTTATCTCTACACC---TGGTCCTGGAGTCTTCAGCAGAAATCTCA 475
          |||| || || | || || |||| || || || || || || || || || || || ||
Db      480 TGATGCCTCTCCTCCTGTACATCTACTCCAGGGGGATCTATGATGGGGACCTGAAGGACA 539

Qy      476 CCATTCCTTATCAGAACATAGGAATTACCCTTGTGTGCCTGACCATTCTGTGGCCTTTG 535
          | || || | || || || || || || || || || || || || || || || || ||
Db      540 AGGTGCCCTATAAAGGCATCGTGATATCACTGGTCCTGGTTCTCATTCCTTGCACCATAG 599

Qy      536 GTGTCTATGTGAATTACAGATGGCCAAAACAATCCAAAATCATCTCAAGATTGGGGCCG 595
          | || | || | || || || || || || || || || || || || || || || ||
Db      600 GGATCGTCCTCAAATCCAAACGGCCACAATACATGCGCTATGTCATCAAGGAGGGATGA 659

Qy      596 TTGTTGGTGGGGTCCTCCTTCTGGTGGTCGCAGTTGCTGGTGTGGTCTGGCGAAAGGAT 655
```

Db	660	TCATCATTCCTCTGTGTCAGTGTGGCCGTCACAGTTCTCTCTGCCATCAATGTGGGGAAGA	719
Qy	656	CTTGGAATTCAGACATCAC-----CCTTCTGACCATCAGTTTCATCTTTCCTTTGATTG	709
Db	720	GCATCATGTTTGCCATGACACCACTCTTGATTGCCACCTCCTCCCTGATGCCTTTTATTG	779
Qy	710	GCCATGTCACGGGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGGA	769
Db	780	GCTTCTGCTGGGTTATGTTCTCTGCTCTCTTCTGCCTCAATGGACGGTGCAGACGCA	839
Qy	770	CAATTTCTTAGAACTGGAGCTCAGAATATTCAGATGTGCATCACCATGCTCCAGTTAT	829
Db	840	CTGTCAGCATGGAGACTGGATGCCAAAATGTCCAACCTCTGTTCCACCATCCTCAATGTGG	899
Qy	830	CTTTCACTGCTGAGCACTTGGTCCAGATGTTGAGTTTCCCACTGGCCTATGGACTCTTCC	889
Db	900	CCTTTCCACCTGAAGTCATTGGACCACTTTTCTTCTTTCCCCTCCTCTACATGATTTTCC	959
Qy	890	AGCTGATAGATGGATTTCTTATTGTTGCAGCATATCAGACGTACAAGAGGAGATTGAAGA	949
Db	960	AGCTTGAGAAGGGCTTCTCCTCATTGCCATATTTTGGTGCTATGAGAAATTCAAGACTC	1019
Qy	950	ACAAACATGGAAAAAAGA	967
Db	1020	CCAAGGATAAAACAAAAA	1037

RESULT 8

US-10-288-222A-15

; Sequence 15, Application US/10288222A

; Publication No. US20030119742A1

; GENERAL INFORMATION:

; APPLICANT: Logan, Thomas Joseph

; APPLICANT: Galvin, Katherine

; APPLICANT: Chun, Miyoung

; TITLE OF INVENTION: Methods and Compositions to treat

; TITLE OF INVENTION: Cardiovascular Disease Using 139, 258, 1261, 1486, 2398, 2414, 7660, 8587,

; TITLE OF INVENTION: 10183, 10550, 12680, 17921, 32248, 60489 OR 93804

; FILE REFERENCE: MPI2001-286P1R(M)

; CURRENT APPLICATION NUMBER: US/10/288,222A

; CURRENT FILING DATE: 2002-11-05

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 15

; LENGTH: 1580

; TYPE: DNA

; ORGANISM: Homo Sapien

US-10-288-222A-15

Query Match 15.3%; Score 173.6; DB 14; Length 1580;

Best Local Similarity 51.9%; Pred. No. 7.2e-46;

Matches 445; Conservative 0; Mismatches 404; Indels 9; Gaps 2;

Qy	119	TGATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGT	178

Db 180 TCATGTTGTTCTTCATCATGCTCTCGTGGGCTGCACCATGGAGTTCAGCAAGATCAAGG 239
Qy 179 CGCACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTG GGGCTCATGC 238
| | | | | | | | | | | | | | | | |
Db 240 CTCACTTATGGAAGCC TAAAGGGCTGGCCATCGCCCTGGTGGCACAGTATGGCATCATGC 299
Qy 239 CTTTTACAGCTTATCTCCTGGCCATTAGCTTTTTCTCTGAAGCCAGTCCAAGCTATTGCTG 298
| | | | | | | | | | | | | | | | |
Db 300 CCCTCACGGCCTTTGTGCTGGGCAAGTCTTCCGGCTGAAGAACATTGAGGC ACTGGCCA 359
Qy 299 TTCTCATCATGGGCTGCTGCCCGGGGGGCACCATCTCTAACATTTT ACCTTCTGGGTTG 358
| | | | | | | | | | | | | | | | |
Db 360 TCTTGGTCTGTGGCTGCTCACCTGGAGGGAACCTGTCCAATGTCTTC AGTCTGGCCATGA 419
Qy 359 ATGGAGATATGGATCTCAGCATCAGTATGACAACCTGTTCCACCGTGGCCGCCCTGGGAA 418
| | | | | | | | | | | | | | | | |
Db 420 AGGGGGACATGAACCTCAGCATTGTGATGACCACCTGCTCCACCTTCTGTGCCCTTGGCA 479
Qy 419 TGATGCCACTCTGCATTTATCTCTACACC---TGGTCCTGGAGTCTTCAGCAGAATCTCA 475
| | | | | | | | | | | | | | | | |
Db 480 TGATGCCTCTCCTCCTGTACATCTACTCCAGGGGGATCTATGATGGGGACCTGAAGGACA 539
Qy 476 CCATTCCCTTATCAGAACATAGGAATTACCCTTGTTGCTGCTGACCATTCTGTGGCCTTTG 535
| | | | | | | | | | | | | | | | |
Db 540 AAGTGCCCTATAAAGGCATCGTGATATCACTGGTCCTGGTTCTCATTCTTG ACCCATAG 599
Qy 536 GTGTCTATGTGAATTACAGATGGCCAAAACAATCCAAAATCATTCTCAAGATTGGGGCCG 595
| | | | | | | | | | | | | | | | |
Db 600 GGATCGTCCTCAAATCCAAACGGCCACAATACATGCGCTATGTCATCAAGGGAGGGATGA 659
Qy 596 TTGTTGGTGGGGTCCTCCTTCTGGTGGTCGCAGTTGCTGGTGTGGT CCTGGCGAAAGGAT 655
| | | | | | | | | | | | | | | | |
Db 660 TCATCATTTCTCTTGTGCAGTGTGGCCGTACAGTTCTCTCTGCCATCAATGTGGGGAAGA 719
Qy 656 CTTGGAATTCAGACATCAC-----CCTTCTGACCATCAGTTTCATCTTTTCTTTGATTG 709
| | | | | | | | | | | | | | | | |
Db 720 GCATCATGTTTGCCATGACACCACTCTTGATTGCCACCTCCTCCCTGATGCCTTTTATTG 779
Qy 710 GCCATGTCACGGGTTTTCTGCTGGCACTTTTTACCCACCAGTCTTGGCAAAGGTGCAGGA 769
| | | | | | | | | | | | | | | | |
Db 780 GCTTTCTGCTGGGTTATGTTCTCTCTGCTCTCTTCTGCCTCAATGGACGGTGCAGACGCA 839
Qy 770 CAATTTCCCTTAGAAACTGGAGCTCAGAATATTCAGATGTGCATCACCATGCTCCAGTTAT 829
| | | | | | | | | | | | | | | | |
Db 840 CTGTCAGCATGGAGACTGGATGCCAAAATGTCCAACCTCTGTTCCACCATCTCAATGTGG 899
Qy 830 CTTTCACTGCTGAGCACTTGGTCCAGATGTTGAGTTTCCC ACTGGCCTATGGACTCTTCC 889
| | | | | | | | | | | | | | | | |
Db 900 CCTTTCCACCTGAAGTCATTGGACCACTTTTCTTCTTCCCCCTCTCTACATGATTTTCC 959
Qy 890 AGCTGATAGATGGATT'TCTTATTGTTGCAGCATATCAGACGTACAAGAGGAGATTGAAGA 949
| | | | | | | | | | | | | | | | |
Db 960 AGCTTGGAGAAGGGCTTCTCCTCATTGCCATATTTTGGTGCATATGAGAAATTCAAGACTC 1019
Qy 950 ACAACATGGAAAAAAGA 967
| | | | | | | | | | |
Db 1020 CCAAGGATAAAACAAAAA 1037

RESULT 9

US-10-085-198-113

; Sequence 113, Application US/10085198

; Publication No. US20040009907A1

; GENERAL INFORMATION:

; APPLICANT: Alsobrook et al.

; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same

; FILE REFERENCE: 21402-279

; CURRENT APPLICATION NUMBER: US/10/085,198

; CURRENT FILING DATE: 2002-02-25

; PRIOR APPLICATION NUMBER: 60/271,646

; PRIOR FILING DATE: 2001-02-26

; PRIOR APPLICATION NUMBER: 60/276,401

; PRIOR FILING DATE: 2001-03-16

; PRIOR APPLICATION NUMBER: 60/311,981

; PRIOR FILING DATE: 2001-08-13

; PRIOR APPLICATION NUMBER: 60/312,858

; PRIOR FILING DATE: 2001-08-16

; PRIOR APPLICATION NUMBER: 60/271,840

; PRIOR FILING DATE: 2001-02-27

; PRIOR APPLICATION NUMBER: 60/277,324

; PRIOR FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: 60/286,096

; PRIOR FILING DATE: 2001-04-21

; PRIOR APPLICATION NUMBER: 60/299,695

; PRIOR FILING DATE: 2001-06-20

; PRIOR APPLICATION NUMBER: 60/315,614

; PRIOR FILING DATE: 2001-08-29

; PRIOR APPLICATION NUMBER: 60/272,405

; PRIOR FILING DATE: 2001-02-28

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 653

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 113

; LENGTH: 1988

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-085-198-113

Query Match 12.5%; Score 141.4; DB 15; Length 1988;

Best Local Similarity 50.8%; Pred. No. 3.5e-35;

Matches 451; Conservative 0; Mismatches 416; Indels 20; Gaps 4;

Qy 88 CTGGAGCTCGTTTTTCACAGTGGTGTCCACTGTGATGATGGGGCTGCTCATGTTCTCTTTG 147
||| | | | | | | | | | | | | | | | | | | |

Db 379 CTGAACCACGGGCTGAACGTGTTTCGTGGGCGCCGCCCTGTGCATCACCATGCTGGGCCTG 438

Qy 148 GGATGTTCCGTGGAGATCCGGAAGCTGTGGTTCGCACATCAGGAGACCC'TGGGGCATTGCT 207
|| | | | | | | | | | | | | | | | | | | | |

Db 439 GGCTGCACGGTGGACGTGAACCACTTCGGGGCGCACGTCCGTCCGGCCCGTGGGCGCGCTG 498

Qy 208 GTGGGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGCTTATCTCCTGGCCATTAGC 267
||| | | | | | | | | | | | | | | | | | | |

Db 499 CTGGCAGCGCTCTGCCAGTTCGGCCTCCTGCCGCTGCTGGCCTTCTGCTGGCCCTCGCC 558

; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES
USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 31375
; LENGTH: 360
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL157789.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.2
; OTHER INFORMATION: NT HIT: gill1435250, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: Q14973, EVALUE 7.00e-64
; OTHER INFORMATION: EST_HUMAN HIT: W01479.1, EVALUE 0.00e+00
US-09-864-761-31375

Query Match 7.0%; Score 79.8; DB 9; Length 360;
Best Local Similarity 56.8%; Pred. No. 2.7e-15;
Matches 147; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

```
Qy      119 TGATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGT 178
          | ||| || | |||| |||| |||| || || |||| || | ||| | |
Db      263 TCATGTTGTTCTTCATCATGCTCTCGCTGGGCTGCACCATGGAGTTCAGCAAGATCAAGG 204

Qy      179 CGCACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGC 238
          | ||| | ||| || || || || || || || || || || || || || || || ||
Db      203 CTCATTATGGAAGCCTAAAGGGCTGGCCATCGCCCTGGTGGCACAGTATGGCATCATGC 144

Qy      239 CTTTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTG 298
          | | || || | | || || || || || || || || || || || || || || ||
Db      143 CCCTCACGGCCTTTGTGCTGGGCAAGGTCTTCCGGCTGAAGAACATTGAGGCACTGGCCA 84

Qy      299 TTCTCATCATGGGCTGCTGCCCCGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTG 358
          | | | || |||| || || || || || || || || || || || || || || ||
Db      83 TCTTGGTCTGTGGCTGCTCACCTGGAGGGAACCTGTCCAATGTCTTCAGTCTGGCCATGA 24

Qy      359 ATGGAGATATGGATCTCAG 377
          | || || || | ||||
Db      23 AGGGGGACATGAACCTCAG 5
```

RESULT 11

US-09-864-761-14847/c

; Sequence 14847, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES
USEFUL FOR

; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

; FILE REFERENCE: Aeomica-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 14847
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL157789.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.2
US-09-864-761-14847

```

```

Query Match          7.0%; Score 79.8; DB 9; Length 560;
Best Local Similarity 56.8%; Pred. No. 3.5e-15;
Matches 147; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

```

```

Qy      119 TGATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGT 178
        | ||| || | |||| |||| |||| || || |||| || | ||| | |
Db      403 TCATGTTGTTCTTCATCATGCTCTCGCTGGGCTGCACCATGGAGTTCAGCAAGATCAAGG 344

Qy      179 CGCACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGC 238
        | ||| | ||| || || || || || || || || || || || || ||
Db      343 CTCATTATGGAAGCCTAAAGGGCTGGCCATCGCCCTGGTGGCACAGTATGGCATCATGC 284

Qy      239 CTTTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTG 298
        | | || || | || |||| || || || || || || || || || ||
Db      283 CCCTCACGGCCTTTGTGCTGGGCAAGGTCTTCCGGCTGAAGAACATTGAGGCACTGGCCA 224

Qy      299 TTCTCATCATGGGCTGCTGCCCCGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTG 358
        | | || || |||| || || || || || || || || || || ||
Db      223 TCTTGGTCTGTGGCTGCTCACCTGGAGGGAACCTGTCCAATGTCTTCAGTCTGGCCATGA 164

Qy      359 ATGGAGATATGGATCTCAG 377
        | || || || | ||||
Db      163 AGGGGGACATGAACCTCAG 145

```

```

RESULT 12
US-09-833-381-317/c

```

```
; Sequence 317, Application US/09833381
; Patent No. US20020132090A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. US20020132090A1el Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 317
; LENGTH: 310
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(310)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-381-317
```

```
Query Match          6.9%; Score 77.8; DB 9; Length 310;
Best Local Similarity 60.2%; Pred. No. 1.1e-14;
Matches 127; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
```

```
Qy      748 CAGTCTTGCAAAGGTGCAGGACAATTTTCCTTAGAAACTGGAGCTCAGAATATTCAGATG 807
      || | || | ||||| | || | ||| ||||| |||| | ||| ||
Db      221 CAACCTGGTACAGGTGCCGAACAGTAGCCTTGGAAACTGGAATGCAGAACTCAGCTG 162

Qy      808 TGCATCACCATGCTCCAGTTATCTTTCACTGCTGAGCACTTGGTCCAGATGTTGAGTTTC 867
      ||| ||||| | || | || || | | || | | | || | ||| ||
Db      161 TGCTCCACCATTGTACAGCTCTCCTTCTCCCCGAGGATCTCAACCTGGTGTTCACCTTC 102

Qy      868 CCACTGGCCTATGGACTCTTCCAGCTGATAGATGGATTTCTTATTGTTGCAGCATATCAG 927
      |||| | || | ||||| | || | || | || | ||
Db      101 CCACTCATCTATACTGTTTTCAGCTCGTCTTTGCAGCAGTNATATTAGGNATTTATGTC 42

Qy      928 ACGTACAAGAGGAGATTGAAGAACAAACATG 958
      || |||| || | | || || ||
Db      41 ACATACAGGAAATGTTATGGAAAAAATGATG 11
```

RESULT 13

US-09-960-352-2253

```
; Sequence 2253, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH
LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
```

```
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 2253
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (390)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 10-LIB34-014-Q1-E1-C5
US-09-960-352-2253
```

```
Query Match          6.5%; Score 74; DB 9; Length 401;
Best Local Similarity 60.1%; Pred. No. 2.3e-13;
Matches 122; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
```

```
Qy      119 TGATGATGGGGCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGT 178
      | ||| ||      |||| |||| |||| || || |||| || || ||| |
Db      199 TCATGCTGTTAACCATCATGCTCTCGCTGGGTTGCACCATGGAGTTCAGCAAGATCAAGG 258

Qy      179 CGCACATCAGGAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGC 238
      |||| || |||| ||| ||| || || || || || || || |||| || ||||
Db      259 CGCACTTCTGGAGGCCCAAGGGGCTGGCCGTCGCTCTGGTGGCGCAGTTTGGCATCATGC 318

Qy      239 CTTTTACAGCTTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTG 298
      | | || || | | |||| || || || || || || || || ||
Db      319 CCCTCACTGCCTTTGGACTGGGCAAGTTCTTCCAGCTGAATAACGTTGAGGCCCTAGCCA 378

Qy      299 TTTCATCATGGGCTGCTGCCCCG 321
      | || |||      |||| |||
Db      379 TCCTGATCTGCNGCTGCTCACCG 401
```

```
RESULT 14
US-09-738-626-2554
; Sequence 2554, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
```

; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 2554
; LENGTH: 972
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-2554

Query Match 6.0%; Score 67.6; DB 9; Length 972;
Best Local Similarity 51.3%; Pred. No. 5.1e-11;
Matches 157; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

```
Qy      129 GCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCGCACATCAG 188
          | | ||||| | | || | | | | | | | | | |
Db      141 GATCATCATGTTACCATGGGTTGACCTTGACGGTGCCCGATTTTCAGATGGTGCTTAA 200

Qy      189 GAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGC 248
          | || | | | | | | | | | | | | | | |
Db      201 ACGTCCACTGCCTATCTTGATCGGTGTAGTAGCGCAGTTTGTTCATCATGCCATTCCTGGC 260

Qy      249 TTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCAT 308
          | | | | | | | | | | | | | | | |
Db      261 GATCGTGGTTGCGAAAATGTTCAACCTCAACCCAGCACTCGCCGTTGGCCTTCTCATGCT 320

Qy      309 GGGCTGCTGCCCGGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGATGGAGATAT 368
          ||| | | |||| | |||| | | | | | | | |
Db      321 GGGATCCGTTCCGGGTGGCACCTCTCCAATGTGATTGCGTTTCTCGCCCGAGGAGATGT 380

Qy      369 GGATCTCAGCATCAGTATGACAACTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACT 428
          | || | || | |||| | | |||| | | || | || |
Db      381 CGCGCTATCGGTCACCATGACCTCTGTGTCCACCATTGTTTCCCCAATCATGACGCCTTT 440

Qy      429 CTGCAT 434
          | ||
Db      441 CCTCAT 446
```

RESULT 15

US-09-738-626-1/c
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626

; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Query Match 6.0%; Score 67.6; DB 9; Length 3309400;
Best Local Similarity 51.3%; Pred. No. 6.1e-09;
Matches 157; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

Qy 129 GCTGCTCATGTTCTCTTTGGGATGTTCCGTGGAGATCCGGAAGCTGTGGTCGCACATCAG 188
| | | | | | | | | | | | | | | | | | | | | |
Db 2466869 GATCATCATGTTACCATGGGTTTGACCTTGACGGTGCCGATTTTCAGATGGTGCTTAA
2466810

Qy 189 GAGACCCTGGGGCATTGCTGTGGGACTGCTCTGCCAGTTTGGGCTCATGCCTTTTACAGC 248
| | | | | | | | | | | | | | | | | | | | | |
Db 2466809 ACGTCCACTGCCTATCTTGATCGGTGTAGTAGCGCAGTTTGTTCATCATGCCATTCTTGGC
2466750

Qy 249 TTATCTCCTGGCCATTAGCTTTTCTCTGAAGCCAGTCCAAGCTATTGCTGTTCTCATCAT 308
| | | | | | | | | | | | | | | | | | | | | |
Db 2466749 GATCGTGTTGCGAAAATGTTCAACCTCAACCCAGCACTCGCCGTTGGCCTTCTCATGCT
2466690

Qy 309 GGGCTGCTGCCCCGGGGGGGCACCATCTCTAACATTTTCACCTTCTGGGTTGATGGAGATAT 368
| | | | | | | | | | | | | | | | | | | | | |
Db 2466689 GGGATCCGTTCCGGGTGGCACCTCCTCCAATGTGATTGCGTTTCTCGCCCCGAGGAGATGT
2466630

Qy 369 GGATCTCAGCATCAGTATGACAACTGTTCCACCGTGGCCGCCCTGGGAATGATGCCACT 428
| | | | | | | | | | | | | | | | | | | | | |
Db 2466629 CGCGCTATCGGTCACCATGACCTCTGTGTCCACCATTGTTTCCCAATCATGACGCCTTT
2466570

Qy 429 CTGCAT 434
| | | |
Db 2466569 CCTCAT 2466564

Search completed: March 25, 2004, 19:04:12
Job time : 505 secs